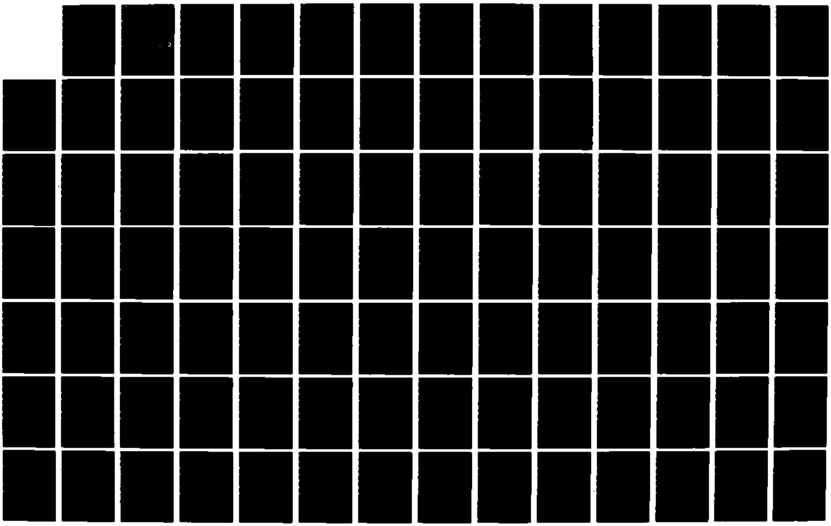
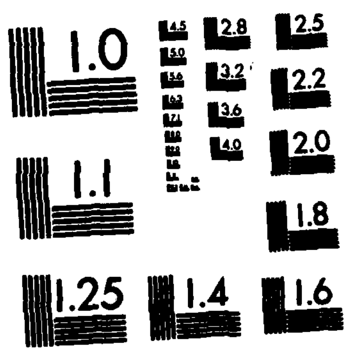
AD-A124 616 COMPONENTS OF MENTAL IMAGERY REPRESENTATION(U) BRANDEIS UNIV WALTHAM MA S M KOSSLYN ET AL. 24 JAN 83 TR-1 N00014-79-C-0982 1/2

UNCLASSIFIED F/G 5/10 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

ADA124616

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: Technical report #1	2. GOVT ACCESSION NO.: AD-A124 616	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Components of mental imagery representation		5. TYPE OF REPORT & PERIOD COVERED: Technical report
		6. PERFORMING ORG. REPORT NUMBER: ONR
7. AUTHOR(s): S. M. Kosslyn, J. L. Brunn, K. R. Cave and R. W. Wallach		8. CONTRACT OR GRANT NUMBER(s): N00014-79-C-0982
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Brandeis University, 415 South Street, Waltham, MA 02254		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR# 150-442, RR# 042-04
11. CONTROLLING OFFICE NAME AND ADDRESS: Personnel and Training Research Programs, Office of Naval Research (Code 458), Arlington, VA 22217		12. REPORT DATE: 24 January 1983
		13. NUMBER OF PAGES: 80
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report): Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

DTIC ELECTE FEB 1 5 1983 S D E

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Final stages of data analysis and report preparation were supported by Contract N00014-82-C-0166

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Individual differences
Mental imagery
Visual thinking
Spatial reasoning
Cognitive components

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The key question asked here was, Is mental imagery ability an undifferentiated general skill, or is it composed of a number of distinct subabilities? Further, if imagery is not an undifferentiated general ability, can its structure be understood in terms of the processing components posited by the Kosslyn & Shwartz theory of imagery representation? A set of tasks was administered to a group of 50 people, and a model was specified for each task. These models invoked different combinations of the processing

components posited by the theory. Each subject was assigned a z score on each measure. The z scores from the different tasks were then correlated, revealing a very wide range of coefficients--which suggests that the subjects were not simply good or poor at imagery in general. In addition, the similarity of each pair of processing models was computed by considering the number of common processing components posited by the theory. The correlations among z scores were then compared to the predicted similarities in processing, and were found to be highly related to these measures. This result suggests that the z scores for task performance in part reflected the efficiency of the underlying processing components, and that for a given person tasks sharing more components tended to be similar in difficulty. Thus, imagery ability is not an undifferentiated general skill, and the underlying components bear a strong correspondence to those posited by the theory. Various additional analyses, considering alternative conceptions and different ways of treating the data, supported these conclusions.

COMPONENTS OF MENTAL IMAGERY REPRESENTATION

Stephen M. Kosslyn
Brandeis University

Jennifer L. Brunn
Harvard University

Kyle R. Cave
Harvard University

Roger W. Wallach
Harvard University

Running head: Components of imagery

Abstract

The key question asked here was, is mental imagery ability an undifferentiated general skill or is it composed of a number of relatively distinct subabilities? Further, if imagery is not an undifferentiated general ability, can its structure be understood in terms of the processing components posited by the Kosslyn and Shwartz theory of imagery representation? A set of tasks was administered to a large group of people, and a model was specified for each task. These models invoked different combinations of the processing components posited in the general theory. Fifty people were tested on the tasks, and each subject was assigned a z score on each measure. The z scores from the different tasks were then correlated, revealing a very wide range of coefficients—which suggests that the subjects were not simply good or poor at imagery in general. In addition, the similarity of each pair of processing models was computed by considering the number of common processing components posited by the theory. The correlations among z scores were then compared to the predicted similarities in processing, and were found to be highly related to these measures. This result suggests that the z scores for task performance in part reflected the efficiency of the underlying processing components, and that for a given person tasks sharing more components tended to be similar in difficulty. Thus, imagery ability is not an undifferentiated general skill, and the underlying components bear a strong correspondence to those posited by the theory. Various additional analyses supported these conclusions.

Components of Mental Imagery Representation

It has long been believed that people differ in their abilities to use mental imagery. In fact, the earliest scientific investigations of mental imagery, reported by Fechner (1860) and Galton (1883), provide support for this notion. The long history of the field notwithstanding, however, little progress has been made in characterizing such differences. In this paper we consider one of the most basic issues in the field, the possible differentiation of "imagery ability" into distinct subabilities.

The attempts to study individual aspects of imagery ability to date have focused on the surface phenomena per se. For example, Betts (1910) and Marks (1974) have devised questionnaires to study individual differences in image vividness and Gordon (1949) devised a questionnaire to study individual differences in imagery control. However, these instruments have had only mixed success in predicting performance (see Marks, 1977; White, Sheehan & Ashton, 1977, for reviews). One problem may lie in the constructs themselves: Comparing aspects of the phenomenological experience is rather like comparing the observable properties of rocks, ice and water: Commonalities in color, shape, rigidity and so on may be misleading, and only by having a theory of structure can one penetrate beneath the surface phenomena themselves. The Kosslyn and Shwartz theory (see Kosslyn, 1980, 1981) provides an analysis of the components underlying visual mental imagery, and the present study examines the usefulness of this analysis as a characterization of individual differences in imagery ability.

The present study, then, has two foci: First, we want to know whether imagery ability is a relatively general, undifferentiated capacity, or whether "imagery ability" is in fact a collection of separate abilities which can vary relatively independently. Second, we want to know whether the components specified by the Kosslyn and Shwartz imagery theory have psychological validity. Insofar as underlying imagery abilities (should they exist) are easily understood as reflecting variations in the efficacy of the components posited by the theory, the theory itself attains additional credibility.

The logic of this investigation is straightforward: We test a group of people on a set of tasks. Each person is assigned a z score for each task. If imagery ability is general and undifferentiated, we expect that people who do relatively well on one task should do relatively well on the others, and people who do poorly on one task should do poorly on the others. Thus, if imagery ability is general and undifferentiated, then we expect the z scores from the various tasks to be highly correlated across the board. At the other extreme, if each task taps a distinct independent "skill," then we expect zero correlation among the z scores. In contrast, if tasks are accomplished using some combination of a small number of available processing components, then the correlation between any two tasks will depend on the number of shared components--with higher numbers shared being reflected by higher correlations. In this last case, then, we expect a wide range of correlations (not just high ones or zero ones), and we expect that the pattern of correlations will reflect the similarity of underlying processing.

Thus, we will first describe the general imagery theory, from which we will derive models for each of our tasks. We next will consider each task in turn, along with the specific model for that task. Finally, we will compare our predictions with the observed correlations among task performance. In this last section we also will compare our theory with a plausible alternative conception.

I. OVERVIEW OF THE IMAGERY COMPONENTS

The components posited by the theory are of two types, <u>structures</u> and <u>processes</u>. Kosslyn (1981) presents a summary of the theory and a computer model embodying the theory, and we will not duplicate his efforts here. Rather, we will highlight the most important structures and processes posited by the theory with an eye toward investigating individual differences in the efficacy of their operation.

<u>Structures</u>

There are two types of structures, <u>media</u> and <u>data structures</u>. A medium does not convey any information inherently; rather, it is a structure that can support data structures. The data structures actually store the information being conveyed. A blackboard or wax tablet are media, which support data structures composed of chalk marks and etchings, respectively. The imagery theory posits distinct short-term memory and long-term memory structures, as noted below:

<u>Short-term memory structures</u>: The "visual buffer" is a special visual short-term memory medium. This structure mimics a coordinate space in the way that an array in a computer can mimic such a space. In a computer, cells in an array correspond to "words" in memory, but the physical arrangement of the words is not like an array. Rather, the words are accessed such that it makes sense to talk about some cells being "adjacent" to or "diagonal" from each other. Similarly, there need be no physical array in the brain in order to have a functional coordinate space, all that is necessary is that representations be organized spatially in the process of accessing them (see Kosslyn, 1980, 1981).

The visual buffer supports data structures that depict information. That is, portions of the medium are activated, and each activated portion corresponds to a portion of the depicted object such that the distances between the portions in the medium mirror the actual distances between the corresponding portions of the object. This depictive data structure is the representation that produces the experience of "having an image," and will be referred to as "the image" in the remainder of this paper. We assume that the visual buffer is used to support images derived from memory and from the eyes during perception proper.

Properties of the medium are especially important because they affect all data structures, all images, that occur within it. As on a TV screen (which is a medium), images begin to fade as soon as they are placed in the medium, the grain of the medium constrains how small an image can be and still be "visible," and the size of the medium constrains how large an image can be while all of it still remains "visible" at once. These last two properties, resolution and extent, place important bottlenecks on how effective one's imagery can be, and thus we have devised tasks that rely on them in the present study.

Long-term memory structures: According to our theory, there are (at least) two distinct media in long-term memory, which store different kinds of information. One stores lists of facts about objects, including descriptions about how parts are put together. The other stores encodings of the "literal" appearance of the object (not a description). The important properties of the long-term memory structures emerge in the context of how various processes operate on them, as is discussed below.

Processes

The various structures posited by the theory can be operated on by various processes, as outlined below.

Processes operating on the visual buffer: Once an image is formed in the visual buffer, it can be operated on in three ways. First, images can be regenerated. Recall that images begin to fade (lose activation) as soon as they are formed in the visual buffer. We posit a REGENERATE process that refreshes units one at a time. Because this process takes time to operate, if too many units are present it cannot refresh all of them before any one of them has faded away. The amount of material that can be maintained at once by this process is critical: In most uses of imagery an image must be maintained over time if it is to be operated upon, and if one cannot maintain much information in an image its usefulness will be severely limited. Thus, we have designed tasks in which performance depends on the amount of information a person can maintain in an image at once.

Second, images can be altered in various ways, such as by reorganizing them into new patterns or by rotating them.[1] We posit a set of specific transformations which are stored as separate routines in our computer simulation model, such as ZOOM, PAN (the inverse of ZOOM), TRANSLATE (move), SCAN, and ROTATE. These process are called up and then given specifications about direction and rate. In addition, we posit that people have the ability to reorganize the internal structure of an image, such as would occur when an imaged Necker cube is seen to reverse; this reorganizing is accomplished by a PARSE process, which will be discussed in more detail later. According to our theory, transformations must always be monitored by an inspection process (discussed below), which allows one to know when the image has been correctly adjusted. We have designed tasks that require people to perform different kinds of transformations.

Third, patterns depicted in images can be "inspected" and classified. Our theory posits a FIND process that is an interface between a semantic description and a pattern in the visual buffer. The FIND process categorizes patterns in the buffer as depicting an exemplar of a given class. This inspection process can be more or less efficient, which should be evident not only when subjects are asked to find a pattern in their image but also when they must monitor a transformation process or when they must find patterns in order to know where to put an imaged part during the course of generating multi-part images (as is discussed below). According to our theory, when the FIND process is used to evaluate an image immediately prior to the subject's making a response, it always first evokes the REGENERATE process so that the image is as clear as possible (see Kosslyn, 1980, Chapters 5 and 7); In addition to the FIND process, we posit a RESOLUTION process which assesses the relative clarity of a part of an image (this was required in order to be able to decide whether to "zoom in" or "pan out" when looking for a part of an image). In our computer simulation model, the RESOLUTION routine computes the dot density of an image (images were displayed by filling in cells of an array), and resolution decreased as density increased (and details ran together--all other things being equal). For purposes of the present study, we considered the RESOLUTION process to be a non-semantic interpreter; it does not identify patterns in the visual buffer as depictions of parts or objects, as does the FIND process. The theory assumes that the image is always regenerated or refreshed immediately prior to using the RESOLUTION process to generate a response. Various tasks made use of each process.

Processes operating on the long-term memory structures: The most important processes operating on the long-term memory structures are those used in generating an image. That is, a pattern in the visual buffer (i.e., the image proper) is generated on the basis of information stored in long-term memory. Two component processes a[illegible] posite[illegible] 1) A PICTURE process activates the stored encodings of appear[illegible], creating a pattern in the visual buffer. This process has considerable flexibility in that it can form images at different sizes, locations and orientations (relative to the visual buffer). 2) A PUT process coordinates separate encodings such that they form a single, composite image. The PUT process is necessary to account for the fact that imagery is a creative activity which can produce new combinations of old things (such as an image of the moon with a giant bull's eye painted on it). According to our imagery theory, parts are activated sequentially, such that an image composed of increasingly more parts will require increasingly more time to complete (see Chapters 4 and 6 of Kosslyn, 1980). In generating multi-part images the PUT process uses the FIND process to find the location where the new part belongs, and then uses this information to set the PICTURE process so that the part is imaged in the correct location (see Chapter 5, Kosslyn, 1980). Thus, we wanted to discover whether the factors that affect image inspection also affect the time to place each additional part in an image, as they should if the same FIND process is used in both operations. We also designed a task that required subjects to use descriptive information to coordinate placement of individual parts of an image.

An important consequence of the theory of image generation is the preservation of stored units in the image itself. That is, because parts begin to fade as soon as they are imaged, each part will be at a different level of activation in the image; this results in the parts being maintained as separate units in the image (because each portion of the unit--each point used to depict it in our computer simulation model--will group together according to the Gestalt law of common fate, with the similarity in level of activation producing the grouping). One of our tasks makes use of this phenomenon, requiring people to use the PARSE process to reorganize the structure of their images.

Finally, there is one last process that is related to the PICTURE process, but which processes input from the eyes rather than long-term memory. The LOAD process retains the contents of the visual buffer that are registering input from the eyes. In retaining this input, additional visual input (which normally supplants the contents of the buffer) is temporarily squelched. Thus, this process is analogous to PICTURE, except that it loads the visual buffer with input from the eyes rather than from information stored in long-term memory.[2]

Formulating Specific Models

Specific models for each of the tasks were formulated within the constraints imposed by our general theory of imagery representation and processing. We assumed that all tasks must be performed using some combination of the structures and processes posited by the theory, and we respected the specific rules of combination posited by the theory (e.g., all images must be formed either via the PICTURE process, from memory, or the LOAD process, from the eyes; the REGENERATE process must be used immediately prior to using the FIND or RESOLUTION process to produce a response; the FIND or RESOLUTION process must be utilized to monitor the progress of any image transformation). In addition, the tasks were designed in such a way that specific components were logically necessary (e.g., using the ROTATE process to rotate an image, using the REGENERATE process to maintain an image over time). Finally, if more than one strategy could be devised, we selected (for better or worse) the simplest one.

The foregoing considerations allowed us to specify which processes should be used in a given task; however, they were not sufficient to determine uniquely the flow among the components. There often are options about the order in which to use various processes (e.g., when to regenerate the image), and it seemed impossible to define tasks in such a way as to preclude such options. Thus, in all analyses in this paper we take seriously only the claims about which components (structures or processes) are recruited in performing a given task, and ignore the details of the order of execution.

All else being equal, the sheer number of shared components should predict the similarity in performance among tasks. Unfortunately, the situation is not so simple because not all components are of equal importance in performing a task. This fact is unavoidable if only because any given measure of task performance is especially sensitive to the efficiency of particular components. That is, measures of the time, speed or correlation across conditions, accuracy, and various ratings data could be collected for each of the tasks, and each of these measures is sensitive to different aspects of processing. Thus, we needed to identify which processes contributed most highly to the performance measure used in each of our tasks. These components were then weighted in subsequent analyses because we expected that individual differences in the efficiency of using these processes would greatly affect task performance.

How could we determine which components had disproportionate influence on a given performance measure? Our method rested on the observation that a given task could be made more or less difficult, depending on the precise stimulus conditions (e.g., images of more complex stimuli are more difficult to maintain). We reasoned that the components that disproportionately affected performance in different stimulus conditions were those that the performance measure was especially sensitive to, and hence were those which would disproportionately contribute to individual differences in task performance. We used two steps to identify these to-be-weighted processes: First, we identified which variations in stimulus conditions in a task (e.g., differences in stimulus complexity) should affect our performance measure. This was done in large part in the light of hindsight, having the results of similar previous tasks in hand. In addition, however, the experiments were designed so that we could verify that specific stimulus factors do in fact affect general performance by examining the group results. Second, we identified how these variations in stimulus conditions affected specific processing components in the task model. This was accomplished by considering the particular roles assigned to the various structures and processes by the general theory outlined above (as will be discussed for each individual task). The components that were responsible for differences in the performance measure in different stimulus conditions were then weighted. If this weighting procedure is amiss, it should become obvious in our later analyses using the individual models.

II. THE TASK BATTERY

General Method and Procedure

The tasks we used all either had been shown previously to tap imagery or were similar to such tasks. Performance on these tasks typically demonstrates some distinctive hallmark of imagery processing, such as increasing amounts of time as the image is rotated or scanned further. By replicating previous results, when group data are considered, we have *prima facie* evidence that imagery was used in performing the task. This is a great advantage over previous studies of separate imagery abilities, which typically provide no evidence that the tests or questionnaires are measuring some aspect of imagery per se. Further, such group results will also provide support for our weighting of specific components.

Each subject was tested individually, for a total of approximately six hours. Testing required three sessions, the third being completed within three weeks of the first. Testing was conducted individually by two experimenters, one male and one female; half of the subjects were randomly assigned to be tested by a given experimenter, with the constraint that each experimenter tested approximately equal numbers of males and females. The experimenters were ignorant of the model for each task and of the specific predictions concerning inter-task similarity. The subjects were told only the general purpose of the study and were ignorant of all hypotheses, specific (i.e., pertaining to group results in a given task) and general. The specific method and procedure for each of the experimental tasks are described below, along with the results for the individual tasks. The tasks were presented to all subjects in the same order, with the same written instructions being used for all subjects. The mental rotation, acuity, and extent tasks were administered during the first session; the image capacity and imaging described scenes tasks during the second; and the image generation, reorganization and inspection tasks during the third. In addition, at the end of the third session the subjects were given a questionnaire and paper and pencil test, as is described below. After completing each task each subject was interviewed about his or her performance on the task; some of the points raised during these interviews will be discussed in the appropriate sections below.

Following the descriptions of the individual tasks will be analyses of the data taken as a whole, examining the interrelations among the various performance measures.

Subjects

A total of 50 subjects were tested. The male experimenter tested twenty-five subjects, 12 women and 13 men. These women ranged in age from 17 to 48 years, with a mean age of 29.6 years. The men ranged in age from 18 to 46 years, with a mean age of 29.2. The female experimenter tested the other 25 subjects, 13 women and 12 men. These women ranged in age from 19 to 47, with a mean age of 30.7. The men ranged in age from 18 to 36, with a mean age of 26.8. All subjects were volunteers, most having answered a classified advertisement placed in the Boston Sunday Globe. The subjects had quite varied backgrounds, as assessed by questions included in an informed consent form administered at the beginning of the first day of testing.[3] Subjects were paid a total of $20 over the three sessions, $4 after the first session, $4 after the second session, and $12 after the third session. This payment schedule encouraged the subjects to complete the entire series, and in fact all but six of the subjects who initially began the study completed all the sessions.

Acuity and Extent

These tasks were designed such that the primary determinants of performance, the weighted components, were either the acuity of a person's imagery or the largest extent at which a person could form an image. The acuity task was a variation of that described by Pennington and Kosslyn (Note 1), and the extent task was a variation of one reported by Finke and Kosslyn (1980).

Acuity Task

In this task subjects imaged a striped grating as if it were moving away from them, and indicated the apparent distance at which the grating seemed to blur. A "blur" criterion was set ahead of time, and subjects were given prior training on distance estimation. The apparent distance at which an imaged grating seemed to blur was used as the index of image acuity.

Materials. Stimuli were projected onto the back of a translucent screen. The screen was 30 inches by 30 inches, but white opaque cardboard covered all of it but a 4 inch wide, 3 inch high rectangle in the center of it. The use of a zoom lens allowed projection of the entire slide in this small window. A chinrest was mounted in front of the screen, 28 cm from its center. Four slides were prepared. These slides all consisted of alternating black and white vertical stripes. The width of the bars on the screen varied such that there were 4, 4.5, 7, and 9 pairs/cm of adjacent black and white stripes for the four slides, respectively.

Procedure. Subjects were first trained to recognize a "blur criterion." While seated with his or her head on the chinrest, the 4.5 pairs/cm grating was presented and then blurred to the point where the bars were not distinct (this point was marked on the barrel of the lens of the projector, and the same amount of blur was used for all subjects). After the subject studied this stimulus for 10 seconds, the slide was brought into sharp focus; if the subject complained that he or she could not remember the appearance of the blurred slide in the time allotted, an additional 10 seconds of study time was allowed before focusing the slide. The subject was then asked to adjust the level of blur, using the focus switch on the remote control for the projector, until the stimulus matched the original one. Feedback was given about whether the slide was blurred too much or not enough, and the slide was refocused into sharp relief. The subject repeatedly readjusted the focus until he or she could adjust the blur to the original level twice in a row.

After learning the criterion, subjects then were given practice at estimating distances. The subjects were asked to stand at specific distances from the screen, ranging from 1 to 9 feet, in units of one-half foot. Distances were presented in a random order, and after each trial the subject was shown the correct location. Subjects performed the task until they were within 6 inches of being correct on four successive distances; in the event that the subject was able to estimate the first four distances correctly, however, an additional two trials were required. Subjects were instructed not to use marks on the floor or other contextual cues to help them estimate the distances, but rather to estimate solely on the basis of the perceived distance from the screen itself.

Following the two kinds of training described above subjects participated in the actual experimental task. The subject was seated at the table with his or her head on the chinrest. One of the three test slides was then presented. The subjects studied the slide for 10 seconds, and then were asked to form a mental image of it. The image was to be at the exact size and proportions as the stimulus in front of them. Once a clear and vivid image was formed, the subjects were to image the slide moving away from them, so it appeared to recede into the distance. Subjects reported that the grating seemed to blur as it seemed to move further away. Subjects were asked to image the grating receding to the point where it was blurred to the level of the original out-of-focus slide used to establish the "blur criterion." At this point, the subjects were to estimate the apparent distance of the grating; that is, the subject was asked, if he or she were actually seeing the grating at that apparent size, how far away would it be? This estimate was our dependent measure.

After making the distance estimate, the subjects were then asked to image the grating rotating 45 degrees clockwise, so it now consisted of oblique stripes. The subjects were then asked to tell the experimenter if the image still matched the blur criterion. If a subject said it did not, he or she was asked to adjust the apparent distance of the grating until it did match the criterion, and then was asked to report whether the image had to be moved closer in or further away in making this readjustment. The reported direction of distance adjustment was recorded. This manipulation was included because Pennington and Kosslyn (Note 1) found that subjects in this task evince an "oblique effect" (see Appelle, 1972), wherein oblique gratings are less acute (i.e., seem closer when they blur) than vertical ones; thus, this seemed a potentially strong measure of how well subjects could "see" images.

Subjects performed this task twice with the 4, 7, and 9 pair/cm gratings, first viewing the thickest, thinest and medium widths and then viewing the thickest, medium, and thinest widths.

The Model

Figure 1 illustrates our model for the acuity task. Subjects first use the LOAD process to form an image in the visual buffer, encoding and retaining the appearance of the grating. They then use the REGENERATE process, which is always invoked when an image must be maintained over time. Only a clear image of segments of two stripes need be maintained to perform the task (in fact, pairs of dots are often used as stimuli in this sort of task; see Finke & Kosslyn, 1980). Thus, the different numbers of stripes need not affect our measure of performance and the REGENERATE process is not weighted. Next, the PAN process is used, and the image is (effectively) contracted. The RESOLUTION process is used to monitor the image as it is being transformed, waiting until the edges of the stripes blur. When this happens, the PAN process is stopped and the distance is estimated on the basis of the apparent visual angle subtended by the field of bars (see Kosslyn, 1978). That is, in the simulation model the angle subtended by the region occupied by the image is one index of distance, and we assume that the size of the grating is preserved in the image (albeit not necessarily with clear or complete images of every bar). Further, we assume that subjects learn the rule relating the angle subtended by the image and distance during the distance-estimation training prior to the experiment.

Our measure of performance for this task is the average distance at which the gratings seemed to blur. According to the model, this distance will be determined by the sensitivity of the RESOLUTION process.[4] If this is so, then wider bars should seem to blur at further distances, as was reported by

Pennington and Kosslyn (Note 1). Thus, the sensitivity of the RESOLUTION process will contribute disproportionately to our measure, and this process will be weighted in later analyses. This analysis receives further justification from Pennington and Kosslyn's finding that the actual distance estimated at the point of blur for the gratings was similar to those estimated by another group which actually observed the gratings in the perceptual analogue to the imaging task (and hence did not even use the other components required in the imagery task).

Figure 2 presents our model of the oblique effect task. The image is now in memory at the point of blur, and must be maintained using the REGENERATE process. The image is then rotated 45 degrees; the FIND process must monitor the ROTATION process, stopping it when the image is rotated the correct amount. Once the image is correctly aligned, the RESOLUTION process determines whether the blur is the same as, less than or more than the blur criterion. The RESOLUTION process is weighted here, for the same reasons as in the first task.[5]

INSERT FIGURES 1 AND 2 HERE

Results

Group Results

Although data from all subjects were considered in the individual differences analyses for each of the tasks, in this and all other tasks reported in this paper only data from subjects who could perform the task were analyzed in the group results. These analyses were intended to discover if imagery was in fact used to perform the tasks and if our analysis of the weighted components received support; hence, we could sensibly consider only data that did in fact reflect task performance. Fortunately, only a very few subjects reported simply being unable to perform any given task.

The data from 45 subjects who completed the acuity task were considered in a single analysis of variance, examining the effects of width of bar grating, trial number, experimenter, and sex of subject. Four subjects were not included in the analysis because they reported being unable to form images of the bar grating presented. A fifth subject was excluded from the analysis because he clearly misinterpreted the task (he gave 330.00 as his distance estimate). As is illustrated in Figure 3, subjects increased their distance estimates as the width of the gratings increased, $\underline{F}(2, 62) = 13.38$, $\underline{p} < .0001$. This result replicates that of Pennington and Kosslyn (Note 1), and thus provides support for the inference that these subjects were doing the task as instructed. Further, these results are evidence that the RESOLUTION process was a primary contributor to the measure; if the thinner gratings had been more difficult to encode or imagine panning away from, we would have expected them to blur at closer apparent distances than the thicker gratings. In addition, although no overall difference was found between estimates given in the first and second presentations of the gratings (trials 1 and 2), $\underline{F}(1, 30) = 1.76$, $\underline{p} > .1$, the increase in estimates for the widest grating on the second trial was larger than was the increase for the narrower gratings (as evident in Figure 3); this result was witnessed by a significant interaction between trial and width of gratings, $\underline{F}(2, 62) = 7.71$, $\underline{p} < .01$. Although there was a tendency for subjects tested by the male experimenter to make larger distance estimates, this effect did not reach significance, $\underline{F}(1, 31) = 3.74$, $\underline{p} < .1$. Interactions between grating size and experimenter, $\underline{F}(2, 62) = 2.69$, $\underline{p} < .1$, and between trial, experimenter and sex of subject, $\underline{F}(1, 31) = 3.79$, $\underline{p} < .1$, also approached significance. No overall effect of sex of subject was found, nor did any other interactions between this variable and grating size or trials reach significance, $p > .1$ in all cases.

INSERT FIGURE 3 ABOUT HERE

Individual Differences Measures

The distance estimates from all gratings in both trials were combined and analyzed together; subjects with higher mean estimates received higher z scores. The mean distance estimate was 51.0 inches, with a standard error of the mean of 5.2 inches. The maximum mean distance estimate given for the three gratings was 190.0 inches, the minimum was 8.0 inches. The subjects who could not perform the task were given z scores .01 below the observed minimum.

Of the subjects who could perform the acuity task, only 10 correctly reported on at least two-thirds of the trials that the oblique gratings seemed closer at the point of blur than had the vertical ones. For purposes of later analyses, we computed the proportion of trials on which a subject reported that the grating became blurrier (i.e., had to be moved closer before reaching the blur criterion). The mean proportion was .32, with a standard error of the mean of .05. The maximum proportion was 1.0, and the minimum 0. We assigned a z score depending on how often they experienced the oblique effect; the subjects who did not complete the acuity test were given a z score of .01 below the lowest z score from the other subjects.

Extent Task

Subjects imaged two gratings moving away from each other (one moving to the left, the other to the right) until the bars seemed to blur. The apparent distance between the gratings at the point-of-blur was used as an index of the horizontal extent of a person's "image field." This task was presented immediately after the acuity task in the same session.

Materials. This task made use of the materials used in the previous task, with one modification. Now, a piece of cardboard was placed over the 4 by 3 inch window in the screen. This mask had two circles cut out of it, side by side; the circles were 2.5 cm in diameter and the rims were separated by 1.5 cm. The circles were equidistant from the vertical midline of the mask and were positioned along the horizontal midline of the mask.

Procedure. The subjects began by viewing the original grating used to establish the blur criterion, now projected through the two circles. The subjects were asked to study it until they were familiar with its appearance through the holes. Subjects were told to imagine that the gratings were actually on the circles. Following this, the subjects closed their eyes and formed a mental image of the two circles, illuminated by unpatterned background light. The subjects then placed the index finger of the left hand directly beneath the left circle and the index finger of the right hand directly beneath the right circle. The subjects then opened their eyes and checked the location of their fingers; this was repeated until the subjects could place their fingers correctly under the circles. Following this, a grating was projected through the circles. The subjects studied the pattern for 10 seconds. The slide was then turned off, and the subjects imaged the circles containing the grating within. As soon as the image was as clear and vivid as possible, the subjects were to image the circles moving apart along the horizontal axis, always keeping the circles equidistant from the center. At the same time, the subjects were to trace the movement of their images by following them with their index fingers. The images were to be moved until they matched the blur criterion; the distance between the fingers at this point-of-blur was measured. As before, each grating was presented twice, now with the medium, thickest, thinest in the first block and the thickest, thinest and medium stripes in the second block.

The Model

The model for the extent task is presented in Figure 4. As in Figure 1, the LOAD and REGENERATE processes are used initially. Now, instead of being panned away from, the TRANSLATE process is used to move the imaged disks away from each other, with the RESOLUTION process monitoring the level of blur. The subject moves his or her fingers along where the image would be, and stops both the movement of the image and the fingers when the gratings blur.

Our measure for later analyses is the mean distance between the gratings at the point of blur. Finke and Kosslyn (1980) showed that the further apart two dots were, the further subjects could image the pair of them moving toward the periphery before they became indistinguishable. The distance between the dots will affect only the RESOLUTION process. Given that we are measuring the extent at the point of blur, and we have varied bar width, our weighting method leads us to weight the RESOLUTION process. We assume that the acuity of this process is a primary contributor to our measure here. (Note that this sensitivity in part also reflects the coarseness of the grain of the visual buffer at different locations.)

INSERT FIGURE 4 ABOUT HERE

Group Results

The data from 41 subjects who completed the extent task were considered in a single analysis of variance, again examining the effects of bar grating width, trial number, experimenter and sex of subject. Eight subjects were excluded from the group analysis because they reported being unable to form images of the grating when the gratings were presented within the circles. A ninth subject was excluded from this analysis because he clearly misinterpreted the task (and gave 87.8 inches as his distance estimate).

As is shown in Figure 5, the wider the grating, the further it was moved into the periphery before the subject reported that it matched the blur criterion, $F(2, 50) = 33.22$, $p < .0001$. This result nicely replicated that of Finke and Kosslyn (1980), providing support for the inference that these data do in fact reflect properties of the image medium and for the inference that the sensitivity of the RESOLUTION process is a primary determinant of distance at the point of blur in this task. There was also a difference between trials in reported distance estimated, $F(1, 25) = 10.46$, $p < .01$, with increased distance estimates given on the second trial. Although there was a tendency for the difference in estimated distance to increase between trials for the wide gratings, the interaction between grating and trial did not quite reach significance, $F(2, 50) = 2.82$, $p = .07$. There were no effects of experimenter or sex of subject, $p > .1$ in all cases.

INSERT FIGURE 5 ABOUT HERE

Individual Differences Measures

For the 41 subjects who reported being able to form images, and who understood the task, the mean distance estimate was 7.0 inches (35.2 degrees of visual arc) with a standard error of the mean of 0.6 inches (3.1 degrees). The maximum distance given was 16.0 inches (71.9 degrees), the minimum was 1.2 inches (6.2 degrees). For purposes of later analysis, each subject was assigned a z score, with larger separations being given higher scores. The subjects who could not perform the task, or failed to understand the instructions, were given a z score .01 below the lowest z score obtained from the other subjects.

Discussion

In general, the group results from the two tasks nicely replicated the results of the original experiments: Broader gratings could be imaged at further apparent distances away from one than could narrower gratings, and at greater extents toward the periphery. The two interesting departures from the previous results were as follows: First, practice apparently differentially aided imaging of the broader bar grating in the acuity task. This difference could be due to the fact that the thickest grating was the one presented first, and thus the subjects had no previous grating to use as a reference point in making distance estimates. Being the widest grating, and thus moved "furthest away," it may have been the most difficult to estimate accurately. Having the experience of estimating the other gratings may have resulted in more accurate estimates the second time around. Some subjects did report something of this sort when interviewed after the tasks.

The second departure from previous results was the failure to find a robust "oblique effect;" only 21% of our subjects reported the effect on at least two thirds of the trials, which is a far cry from the overwhelming majority found in the Pennington and Kosslyn (Note 1) study. The use of a blur criterion may have been critical here: The oblique effect is known to be very subtle, and the blur criterion may not have been blurred enough to pick up the slight differences; the Pennington and Kosslyn study did not use such a criterion, but merely asked subjects to image the patterns moving away until they seemed to blur into homogenous gray. In addition, the Pennington and Kosslyn study required subjects to image a grating at a specific orientation initially, and then to estimate the distance at the point of blur. It is possible that the requirement that subjects rotate an image already at the point of blur (in the present experiment) was simply too difficult.

Mental Rotation

In this task subjects were shown alphanumeric characters at different tilts. Half the letters faced normally and half were mirror-reversed. Subjects were asked to image the character revolving in a clockwise direction until it was upright, and then were to classify the direction in which it faced. Cooper (1975) demonstrated that subjects could rotate images in a single direction, and we used her instructions in our effort to obtain more stable estimates of an individual's speed of mental rotation.

Materials

The stimuli were 60 slides of 5 different alphanumeric characters: R, G, 2, 5, and 7. Each character appeared twice at each of six orientations, 0, 60, 120, 180, 240, and 300 degrees rotated clockwise from the standard upright orientation. At each orientation the character appeared once facing normally and once mirror-reversed. The slides were projected on the rear of a translucent screen such that the characters subtended about 7 degrees, as viewed from the subject's position in a chinrest. Two keys were connected to a reaction-time clock that started when a slide was projected and stopped when either key was pressed.

Procedure

The subjects began by reading the instructions, which explained that we were interested in the rotation of visual mental images. The subjects were told that when a character was presented they should imagine it rotating clockwise until the top of the character was at the top of the image. At this point, the subjects should "look" on the image and see which direction the character faced. If the character faced in its normal direction, the subjects were to press the key under their dominant hand; if it faced in the mirror-reversed direction, the subjects were to press the other key. It was stressed that the subjects should always rotate images in the clockwise direction and should always respond only after examining an upright image.

Six practice trials, using the numbers 3 and 4 preceded the actual test trials, and the subject was interviewed about his or her performance during the practice trials; any misconception about the task was corrected. On all trials, the slide remained illuminated until a response was made and there were 10 sec between a response and presentation of a new slide. The stimuli were presented in a random order, except that the same character could not appear twice within three consecutive trials and no more than three trials in a row could present characters at the same orientation or facing in the same direction. Subjects were asked to respond as quickly as possible without making errors while still following the instructions.

The Model

Figure 6 presents our model for the mental rotation task. The LOAD process is used to encode the stimulus, and then the ROTATE process is used to rotate the image clockwise. As the image is rotated it is monitored by the FIND process, which stops the rotation when the figure is upright. After the image is correctly oriented the REGENERATE process is used prior to using the FIND process to classify the direction in which the character faces (which in part involves first identifying the character).

The speed of rotation is our measure, and orientation was varied. Differences in orientation should require different amounts of processing by both the ROTATE and FIND processes; rotation speed is a function of the speed of both processes, assuming that the ROTATE process will not operate faster than the FIND process can monitor the orientation. Thus, these two processes are weighted in later analyses.

INSERT FIGURE 6 ABOUT HERE

Results

Group Results

The data from all 50 subjects were considered in a single analysis of variance, examining the effects of angle of presentation, type of response, experimenter, and sex of subject. First, our results replicated earlier investigations of mental rotation (e.g., Cooper, 1975), with increasingly more time being required to classify figures that were rotated increasing amounts, $F(5, 170) = 20.76$, $p < .0001$. This finding is illustrated in Figure 7 and provides support for the claim that subjects were in fact rotating images as instructed. Although "normal" judgments were made more quickly than "reversed" judgments, $F(1, 34) = 22.85$, $p < .0001$, this difference did not vary for the different orientations, $F(5, 170) = 1.33$, $p > .25$ (even though in Figure 7 the items for stimuli at 300 degrees seem to diverge for the two kinds of judgments). There were significant differences in the time to judge the different items, $F(4, 136) = 2.91$, $p < .05$. But this result seemed due to two interactions: First, the increases in times for different amounts of rotation were less pronounced for some items than for others, with "5" showing the sharpest increase and "G" the shallowest, $F(20, 680) = 1.62$, $p < .05$. Second, the time to classify stimuli as "reversed" varied for the different stimuli, as is evident in Figure 8, $F(4, 136) = 4.00$, $p < .01$. In addition, there was no effect of sex of subject, F 1, and the overall effect of experimenter was not quite significant, $F(1, 34) = 3.34$, $p < .1$. However, time increased less sharply for increased amounts of rotation for subjects tested by the female experimenter, $F(5, 170) = 2.35$, $p < .05$. The only other effect or interaction to attain significance was between item, sex of subject, and experimenter, $\underline{F}(4, 136) = 2.77$, $\underline{p} < .05$; this result reflected the fact that "7" was relatively fast for the male subjects tested by the female experimenter. No other effect or interaction was significant, $\underline{F} < 1.5$ in all cases.

INSERT FIGURES 7 AND 8 ABOUT HERE

There was a mean of 10.9% errors overall. Errors did not tend to increase as times decreased, belying a speed-accuracy tradeoff: Mean error rates were 8.4%, 7.2%, 8.8%, 13.0%, 15.2%, 13.2% for 0, 60, 120, 240, and 300 degrees, respectively.

Individual Differences Measures

Our measure was the speed with which subjects rotate an image. Because there was some indication that some subjects rotated the 300 degree stimuli the "short way around," we computed slopes on times from only those stimuli presented at 240, 180, 120, 60, and 0 degrees. We reasoned that if the slope was not positive (as it was not for three subjects) we could not take the slope to reflect speed of rotation. The mean slope over the five orientations (0 through 240 degrees) was 11.0 msec/deg, with a standard error of the mean of 1.6 msec. The maximum slope was 52.3 msec/deg, and the minimum slope was .06 msec/deg (not including the three subjects with negative correlations; the minimum including all subjects was -1.0 msec/deg.). Mean slopes were converted to $\underline{z}$ scores for later analysis, with shallower slopes being given higher scores. The three subjects producing negative slopes were given $\underline{z}$ scores .01 lower than that obtained for the steepest slope.

Discussion

The results of this experiment nicely replicated the usual findings when the group data were examined. In fact, the group analysis produced an effect of items very much like that reported by Hock and Tromley (1976), who found that rounder letters--such as G--tended to not be rotated as far as other letters, which is reflected by differences in slope in our analyses. Hock and Tromley interpret their results as showing that people only rotate figures until the top of the figure is at the top of the picture-plane. However, in contrast to the usual results, the actual magnitude of the slope found here was very steep: These subjects on average were about three times slower than the well-practiced college students tested by Cooper (1975). The only real surprise in the group analysis was the different results for the two experimenters. We have no ready account for these results, and suspect that they are not worthy of too much concern until replicated.

Line Drawings

The performance measures in this task were designed to be especially sensitive to how much material a given person could maintain in an image at once. Subjects imaged configurations of different numbers of short lines connected end to end, made a speeded judgment about the image, and then drew the configurations. This task was modeled after one originally reported by Bower (1972), who found that the inclusion of more lines impaired memory for the configuration.

Materials

Configurations of lines were constructed, composed of short segments linked to form a pathway. A total of 18 different configurations were used, two each containing 2 through 10 segments. A tape recording was prepared containing the sequence of trials. Each trial consisted of a series of compass directions (e.g., "north, northeast, west") which described how successive line segments should be attached together, end to end. The segments used the eight major points of the compass and were combined in a random order, except that no more than two consecutive segments could point in the exact same direction, no segment could reverse on the one immediately before it, and lines did not cross. All configurations were open-ended. A new direction was presented every four seconds. All compass directions were recorded on only one channel of a stereo tape recorder; four sec after the final compass direction in a sequence the word "endpoint" was recorded on both channels, and four sec after this the word "draw" was recorded on both channels. The subjects heard every stimulus, and the tape recorder was connected to a sound-activated relay such that whenever a word was recorded on both channels a reaction-time clock started and the tape stopped. When the subject pressed either of two telegraph keys, labeled "above" and "below," the clock stopped and the tape resumed. The first nine stimuli each had a different number of line segments, as did the second nine stimuli. Trials were presented in a random order within each block of a stimuli.

Procedure

Subjects were told that they would hear descriptions of how 1-inch-long line segments were attached together, end to end. They should mentally image the path formed by connecting up the segments as if they were viewing a drawing held at arm's length. They were told that the direction from the starting point would be given initially, and then the direction from the tip of the first line would be given, indicating how to image the second line pointing from it and so on. Subjects were given practice drawing, on paper, line drawings similar to the images they would soon be asked to form. They practiced these drawings until they were able to complete one with no mistakes (i.e., all directions were drawn correctly).

Subjects were told that paths would differ in both the number of segments used and in the overall configuration formed. They should try to hold all segments in the correct locations in an image throughout the task. Subjects were asked to close their eyes while imaging the drawings. Upon hearing the word "endpoint" the subjects were to evaluate their images, and to indicate, by pressing the appropriate button as quickly as possible, whether the endpoint was above or below the initial starting point. This task was included partly to motivate the subjects to form an image, rather than simply rehearsing the directions or the like. We also expected, however, that more time would be required to inspect fuzzier images (see Kosslyn, 1975). The answer given and the time to respond were recorded from this task. Next, the subject heard the word "draw," which was a signal to begin drawing the line configuration that was imaged. A 4 by 4 inch area on a sheet of paper was provided for each drawing. Again, the subjects were instructed to work as quickly and accurately as possible. As soon as the subjects were finished drawing they were to press the key used to respond "above," which started the tape recorder and began a new trial. The drawings were sufficiently simple that we did not worry that drawing ability per se would contaminate our measure (and in fact no subjects had difficulty in drawing the initial configurations used in explaining the task). A new trial began six seconds after the subjects indicated that they had finished drawing the previous configuration. Prior to the actual test trials subjects practiced imaging and drawing with 5, 7 and 9 segment configurations.

<u>The Model</u>

Figure 9 presents our model for the drawing task. The subject uses the PICTURE process to image the first line at the center of the visual buffer. The REGENERATE process is used, maintaining the image until the next command is heard. If this command specifies the location of a new item, the PUT process uses the FIND process to locate where the new segment should be placed, and then uses the PICTURE process to image a new segment in the correct location. Once this is done, the image is again regenerated until a new command is received. When the command is not a new direction, but is the word "endpoint" a second task (described below) is performed and then the image is drawn.

The measure used in this task was the mean number of segments correctly drawn. The stimulus conditions vary in the number of segments in the image. When more segments are present, the only component that must do more work when a new segment is added is the REGENERATE process. Thus, this component is weighted in subsequent analyses.

INSERT FIGURES 9 AND 10 ABOUT HERE

Figure 10 presents our model of the task requiring comparison of the relative heights of the endpoints of the imaged path. The REGENERATE process is first used to make the image as clear as possible. Next, the subject "glances" in the rough direction of the start point. If the start point is not in the region of highest acuity in the visual buffer, the subject will use the SCAN process to "see" the start point clearly. We assume that the subject can quickly compute the rough direction in which to scan on the basis of at least partial memories of the directions per se, which were used by the PUT process position the line segments correctly (see Kosslyn, 1981). Finally, the actual spatial judgment is carried out by the FIND process.

Our measure here is the time to make the above/below decisions. The various stimuli differ in the number of segments, which will tax the REGENERATE process to greater or lesser degrees, and the distance between the endpoints which will affect both the time required to scan and the actual decision time. Thus, our method for determining weighting leads us to weight all three components used in performing this task.

Results

Group Results

The data from all 50 subjects were considered in separate analyses of variance examining the drawing scores and probe times. In the analysis of the drawing scores we examined the effects of number of lines in the figure, effects of practice, experimenter and sex of subject. As shown in Figure 11, the subjects were generally less accurate in reproducing the line drawings that contained more segments, as witnessed by decreased drawing scores (percent of lines drawn correctly), $F(8, 272) = 51.36$, $p < .0001$. Subjects improved with practice, as indicated by an overall increase in the drawing scores on the second trial, $F(1, 34) = 9.52$, $p < .01$. The mean drawing score for the first trial was 50.56, for the second trial the mean score was 56.72. However, this increase was not consistent for all drawings, $F(8, 272) = 8.01$, $p < .0001$; as is evident in Figure 11, the decrease in drawing scores with more line segments is not a smooth curve, possibly because some drawings could be more easily organized into relatively few higher-order "chunks" (e.g., several of the drawings had contiguous lines going in the same direction). Drawing scores did not significantly vary with sex of the subject or with experimenter, p .2 in both cases. However, an interaction between experimenter and sex of subject approached significance, $F(1, 34) = 3.36$, $p < .1$, with female subjects tested by the male experimenter having the lowest drawing scores.

In the analysis of variance performed on the probe times we examined the position of the endpoint relative to the starting point (i.e., above or below), the number of lines in the drawing, effects of experimenter, and sex of subject. As shown in Figure 12, there was a difference in the time to respond to probes for the different stimuli, $F(8, 368) = 7.76$, $p < .0001$; however, times did not show consistent change as line length increased. A regression analysis performed on these data revealed that the main determinant of times was the distance between the start and endpoints, with more time taken to judge closer--and hence less discriminable--points, $p < .01$; there were no effects of number of segments per se in this analysis, $p > .25$. These results suggest that subjects gave the first and last segments a high priority in their efforts to maintain their images--perhaps in part because they knew they would need to remember these segments to do this judgment task. The only other significant result in the analysis of variance was an interaction between judgment (above vs. below) and stimulus, $F(8, 368) = 7.55$, $p < .0001$, which also reflected the effects of distance between the endpoint and starting point. No other interaction reached significance, $p > .05$ in all cases, nor were there effects of any other variable, $p > .1$ in both cases.

INSERT FIGURES 11 and 12 ABOUT HERE

There was a mean of 10.8% errors overall on the endpoint judgments. Such relatively accurate performance on this spatial relations task provides evidence that subjects were in fact forming images, rather than using verbal rehearsal.

<u>Individual Differences Measures</u>

We created two measures of performance in the drawing task, and one in the judgment task. The second measure in the drawing task was created merely to help shed light on the results, and was not used in the subsequent analyses. The descriptive statistics for each measure are as follows:

The first measure was the mean drawing score for all the drawings. The drawing score for each drawing was computed using the following formula: c/c+w x 100 (c = number of lines in the drawing, w = number of lines drawn incorrectly or lines added or missing). Drawings were scored by considering each pair of lines in turn. If two lines were correctly positioned relative to each other, they were considered correct--even if previous lines were incorrect. Two judges scored the drawings independently and the few disagreements were resolved easily. The mean drawing score was 53.6 (out of a possible total of 100), with a standard error of the mean of 2.3. The maximum mean drawing score was 82.7, the minimum mean drawing score was 14.3. Higher $\underline{z}$ scores were assigned to subjects with higher drawing scores.

The second measure was the number of lines in the most complex drawing the subject drew correctly both times while also drawing correctly both instances of the line drawing with one less segment. This was a very conservative measure, ensuring that "luck" or peculiarities of stimuli that happen to contain a given number of segments (e.g., symmetrical parts) would not be responsible for a score. The mean number of lines in the most complex drawing was 1.02, with a standard error of the mean of 0.24. The maximum number of lines was 5.00, the minimum was 0.00.

The third measure was the overall mean time subjects took to decide whether the endpoint was above or below the starting point. The mean of these times was 2262 msec, with a standard error of the mean of 225 msec. The maximum mean answer time was 9733 msec, the minimum time was 843 msec. Higher $\underline{z}$ scores were assigned to people having faster times.

Discussion

The most interesting result found in this task was the remarkably poor performance exhibited by most of the people. The mean number of lines in the drawing the subjects drew correctly, and drew correctly both instances of the drawing with one fewer segment, was only 1.02. This result is especially surprising given Weber and Bach's (1969) report that their subjects could hold a mean of six letters in an image at once; their subjects were, however, university students and their materials may have been easier to organize into higher-order "chunks." The other results of interest are the existence of an effect of practice, even with the little practice given here, and the existence of item differences, which seem due to differences in the ways the paths could be organized into higher-order units. These two effects are interesting when taken together: Perhaps practice effects occur because subjects learn to chunk the segments so that, for example, a right-angle elbow comes to be imaged as a single unit rather than two separate segments. These results suggest that it may be possible to train subjects to become better at maintaining more material in their images.

Imaging Described Scenes

This task was designed to be sensitive primarily to a person's abilty to image scenes on the basis of verbal descriptions. Subjects heard the names of four objects and their relative locations; after imaging the described scene, subjects scanned between a given pair of objects and time to scan was measured. Scan time was used as an index of distance, which was used to assess how accurately the scene was constructed. This paradigm is a variation of one described by Kosslyn, Reiser, Farah, and Fliegel (see Chapter 4 of Kosslyn, 1980).

Materials

The items in this experiment were 48 scenes, each of which was composed by placing four objects in specified spatial relations. The objects were named by concrete nouns selected primarily from the Paivio, Yuille, and

Madigan (1968) norms. All of these nouns had a sum rating of Imageability and Concreteness of at least 12.4, precluding concrete nouns such as "dell" which people know are concrete but have trouble imaging or imageable nouns such as "amazement," which do not name actual objects. In addition no nouns were used that named very large objects, were strikingly affective (e.g., "corpse"), would not take an indefinite article (e.g., "blood"), were simply nominalized adjectives (e.g., "sunburn"), were not a "basic level" noun (in Rosch's, 1978, sense--e.g., "limb"), named an action (e.g., "kiss"), were the name of an occupation (e.g., "policeman") or were names of things that were not easily localizable (e.g., "orchestra "). These criteria were required to produce nouns that named easily imaged, single objects. The criteria were so strict, however, that we fell short of selecting the requisite 192 nouns. Thus, we constructed a list of additional candidate items that met the criteria noted above, and gave these 227 items to 12 Harvard undergraduates. These people were asked simply to cross out any noun for which they had any trouble at all in forming an image of the named object. Just over a hundred were not crossed out on any subject's list, and 98 of these were then selected at random. The extra items (along with a few additional words selected by the experimenters) were used in the eight additional scenes created for use in practice trials.

The nouns thus selected were randomly arranged into groups of four. The items in each scene were linked into a description in which the location of each object was specified relative to the previous one in the sequence. The objects were specified with respect to distance (1, 2, 3, or 4 inches) and direction (up, down, left or right). One scene, for example, was "Briefcase, 4 inches up place a Horse, 1 inch left place a Beaver, 1 inch down place an Onion." Scenes were constructed such that at least six scenes incorporated one of eight different distances between specified pairs of objects. The inclusion of these different distances is important for the scanning task used to assess image accuracy, as will be described shortly. We also took care to ensure that no direction would occur more than twice in one scene, a direction and its exact opposite occurred no more than once per scene, and the maximal total distance in one direction was 5 inches (not the theoretically possible 8 inches).

The 48 descriptions were recorded only on the channel heard by the subject, with a two second pause between an object name, the next distance and direction, and the next object name. Two seconds after the name of the fourth object the word "image" was recorded on both channels, starting the reaction-time clock and stopping the tape recorder. A response stopped the clock and started the recorder. Two seconds later the name of one of the objects was repeated, recorded on only one channel, followed four seconds thereafter by the name of another one of the four objects now being recorded on both channels (and thus again starting the clock and stopping the tape). Six seconds later the name of a new object was presented and a new trial began. The two repeated objects constituted the "probe." The serial position of these two objects was balanced over all trials and each position occurred almost equally often in each of the eight-distance conditions (because there were six probe pairs at each of the eight distances but there are only four serial positions per scene, the balancing could not be perfect). In addition, on half the trials for each distance the first object named was to the left of the second and vice versa for the other half. The order of the trials themselves was random with the following constraints: First, each distance occurred three times during the first half of the trials and three times during the second half (in order to control for possible effects of practice), and the first object was to the left or right of the second object equally often on both halves of the trials. In addition, the first item named in a probe could not be on the same side more than three times in a row, and the same distance between probe items could not occur more than twice in a row.

Procedure

The subjects were told that they should image the named objects as though they were drawings on a page held at arm's length, and that these drawings should be arranged in accordance with the directions and distances on the tape. When the word "image" was presented, the subject was told to form a clear image of the entire scene, with the objects arranged as described, and push the button marked "right" when the image was complete (this task uses the basic methodology developed by Beech and Allport, 1978). After indicating that the image was fully formed, the subjects were told that they would hear the name of a "focus" item. The subjects were to "mentally stare" at this item, keeping the spatial arrangements among all the items intact while doing so. Shortly thereafter, the name of a second item would be presented, and the subjects were to image a tiny black speck flying from the first item to the second one in the shortest direct path, moving as quickly as possible (this task was used very successfully by Kosslyn, Ball, and Reiser, 1978, who found a .97 correlation between time to track the speck and distance). As soon as the speck arrived at the second object, the subjects were to press the button marked "right" or the one marked "left," depending on which way the speck flew (ignoring the vertical component). None of these discriminations was difficult. Subjects were told that the speck should be seen as moving as quickly as possible while still remaining visible. A drawing of a scene was then presented, focus and target items were named, and it was made clear that diagonal distances between objects were to be traversed freely by the speck. The subjects were urged to maintain the same scale size from scene to scene and to image objects in the correct spatial relations at all times.

Prior to the actual test trials the subjects were asked to draw training trials at the correct locations and distances; all drawings were done on a piece of paper at the distance at which the subject was to image drawings on the test trials. On these training trials, however, letters of the alphabet were named in order to minimize drawing difficulty. After each scene was completed the experimenter measured the distances drawn by the subject and pointed out errors. This procedure was repeated until the subject could draw a scene with all three distances specified in the scene within 1/4 inch of the correct value. All subjects completed at least four training trials, which included all four distances used in the test trials (the maximum required by any one subject was eight). Following this, a series of eight practice trials presenting names of objects was presented and the subject performed the imagery task. The tape was stopped after at least one trial and the subject queried about his or her performance on that trial; if the subject described doing something incorrectly, the instructions were reiterated and additional practice trials presented (with additional queries about he subject's performance during these trials). Thus, by the time of the actual test trials we had good reason to believe subjects understood how to image the scenes and how to perform the scanning task, giving us some confidence that the correlation between the scan time and the distance that should have been scanned would reflect how accurately the subjects could use the descriptions to form an image of a scene.

The Model

Figure 13 presents our model for the described scenes task. The PICTURE process is used to image the first object named. For the following three items the direction and distance are input to the PUT process, which (via the

FIND process, to locate the previous object, which is used as a "foundation part" for the new one, and the PICTURE process to form the image) places the subsequent images in the correct locations. After each item is arranged, the REGENERATE process is used to maintain the image. After the image is completed, subjects focus on the first probe item, which is located via the FIND process. They then scan to the second object, with the FIND process monitoring the SCAN process so scanning is direct. Finally, when the FIND process discovers that the target item is in focus, a response is issued. The response is determined by the direction, left or right, of the trajectory of the scan path.

Our measure of performance in this task is the correlation between the assumed distance and scan time. This measure is especially sensitive to the variations in the distance among objects--which were produced by the PUT process--and the regularity and direction of the SCAN process. Thus, the PUT and SCAN processes were weighted for later analyses. (Because we are not assessing the speed of scanning, the FIND process is not weighted.)

INSERT FIGURE 13 ABOUT HERE

Results

Group Results

The data from all 50 subjects were considered in a single analysis of variance, examining the effects of distance scanned, direction of scan, practice at scanning a particular distance, experimenter and sex of subject. To our surprise, there was no systematic relationship between scan time and presumed distance $F(7, 238) = 1.27$, $p > .25$. There was no overall difference between time taken to scan a particular distance to the right or to the left,

$\underline{F}(1, 34) = 2.74$, $p > .10$; however, as is evident in Figure 14, differences between the direction of the scan did vary with distance, $\underline{F}(7, 238) = 3.55$, $p < .01$. The effects of practice also varied with distance, $\underline{F}(14, 476) = 2.24$, $p < .01$, with practice seeming to have its greatest effects in speeding up performance on trials in which subjects scanned medium distances. There were no effects of experimenter or sex of subject, $\underline{p} > .3$ in both cases.

INSERT FIGURE 14 ABOUT HERE

There was a mean of 14.5% errors overall. The error rates were 15.7%, 20.0%, 12.0%, 12.0%, 9.7%, 22.7%, 12.7%, and 11.7% for 1, 1.4, 2.24, 3.0, 3.61, 4.24, 5.0, and 5.66 inches respectively. These high error rates indicate that many subjects did not include the correct spatial relations in their images.

Individual Differences Measure

The correlation between the distance of the objects as stated in the description and the time taken to scan between the objects was computed for each subject. The mean correlation was $\underline{r} = .02$, with a standard error of the mean of .06. The maximum correlation was $\underline{r} = .95$, and the minimum was $\underline{r} = -.68$. For purposes of later analysis, each subject received a $\underline{z}$ score on the variable, with higher $\underline{z}$'s reflecting larger correlations.

Discussion

The results from this task were very surprising: There were no reliable increases in scanning time with increases in the distance presumably scanned. The effects of distance on scanning time have been replicated repeatedly (see Beach, 1978; Kosslyn, 1973, 1978; Kosslyn, Ball, & Reiser, 1978; Pinker, 1980; Pinker & Kosslyn, 1978; Pylyshyn, 1981). One explanation for our failure to find effects of distance on scan time is that most subjects simply did not maintain the correct distances in their images. This could have been because they were not able to maintain four items in an image at once (recall the dismally low image maintenance scores we obtained). However, in previous work (e.g., Kosslyn et al., 1978) subjects scanned between up to seven different imaged objects with no trouble; but of course these subjects were all university students. Alternatively, subjects may not have been able to use the descriptions to place images in the correct juxtaposition in an image, even though they could use them to draw correctly. These two accounts may not be independent, of course: It could be that subjects could not hold all the images at once partly because of the effort involved in comprehending the instructions and placing each new image. Yet another account is suggested by the surprisingly low estimates we sometimes obtained in the extent task. Many of our subjects claimed not to be able to image gratings even three inches apart (at approximately the same distance from their eyes as the images formed here), and many of the scenes described here extended over five inches. Thus, subjects may not have been able to fit all objects into an image at correct distances--and hence would not have scanned the distances we used to compute the correlation.

Image Generation, Reorganization, and Inspection

This task had three parts, which allowed us to measure 1) how quickly subjects could add additional parts to an image; 2) how quickly they could reorganize an image into a set of different shapes; and 3) how quickly they could "see" parts of imaged forms. Subjects viewed geometric forms that were described in terms of different numbers of units. The generation task is identical to one described by Kosslyn, Reiser, Farah, and Fliegel (see Chapter 4 of Kosslyn, 1980), where it was shown that image formation times increase with the number of parts delineated by a given description of a form. The inspection task is a variant of one used by Kosslyn et al. It has been shown in this task that more time is required to "see" parts that are composed of segments belonging to different units in the image (e.g., a hexagon is harder to see in an image of the Star of David than is a triangle; see Reed, 1972). Thus, if we replicate the earlier findings here we have prima facie evidence that these subjects were in fact using imagery. Finally, although the reorganization task is novel, our theory makes a simple prediction for the reorganization times, which if confirmed will provide evidence that the subjects used imagery in this task. Namely, it was expected that it would be easier to reorganize images from overlapping to contiguous rather than vice versa. This prediction was based on the notion that parts are maintained as separate units in the image, and hence one must often violate part boundaries when reorganizing images. This sort of cross-boundary reorganization will occur more frequently when one reorganizes relatively many parts into fewer ones. For example, when reorganizing five squares into two overlapping rectangles (see the lower left figure of Figure 15), one will have to use the FIND process to coordinate numerous lines belonging to different parts, which hence are at different levels of activation. This is more difficult than reparsing two overlapping rectangles into five squares; in this case, much of the reparsing does not involve coordinating lines belonging to different units, but dividing up lines which are at the same level of activation (e.g., a side of a rectangle).

Materials

The stimuli were the ten geometric figures illustrated in Figure 15. Each of these figures can be described in two ways, either in terms of overlapping shapes (e.g., for the upper left pattern in Figure 15, a square and two triangles) or in terms of adjacent shapes (four squares and four triangles). In general, more shapes are posited when the figures are described in terms of contiguous shapes. Two booklets were prepared, each containing two instances of each of the ten figures. The booklets differed only in which type of description--overlapping or contiguous shapes--accompanied each figure. The same type of description was used for all figures in a given booklet. The booklets were composed such that in the center of one page was a typewritten description of a figure to follow, and on the next page the figure appeared with the description being presented beneath it. The figures were centered on the pages, all being about the same size, drawn in outline on 8 by 11 inch white paper. The pages were numbered. The figures were randomly ordered except that each figure was presented before any figure was presented twice. The drawings were in the same order in both booklets. In each booklet two instances of each of three practice figures, with accompanying descriptions, were presented before the actual test stimuli.

INSERT FIGURE 15 ABOUT HERE

A tape recording was prepared, with stimuli on it being coordinated with pages in the appropriate booklet. The first stimulus recorded for each trial was a page number, corresponding to the number of a page containing the description of a figure. Four seconds later, the succeeding page number was presented, corresponding to the page containing the figure (with description). Eight seconds thereafter the word "cover" was presented, followed in two seconds by the word "image." The word image was recorded on both channels, thus stopping the tape recorder and starting a reaction-time clock. Four seconds after the tape recorder was started again, one of two stimuli was presented: either the alternative description (adjacent shapes if the initial description was overlapping shapes, or vice versa) or a probe (as will be described shortly). These stimuli were recorded on both channels. If a second description was presented, a probe was presented four seconds after the subject responded. The next consecutive number was presented ten seconds after the probe, marking the beginning of a new trial.

In constructing the trials the figures were divided into two groups of five, with each group having an equal total number of contiguous parts comprising the figures. One group of five figures was used in one booklet in the alternate description condition whereas the other group of five figures was used in the immediate probe condition. The assignment of item groups to conditions was reversed in the other booklet. Thus, each figure was reorganized, but only in one direction (from many to few or vice versa). Similarly, each figure was probed immediately after being imaged, but again only in one organization. There were two types of probe questions asked, either querying the presence of a particular axis of symmetry (vertical, horizontal, or diagonal), or querying the presence of a particular geometric shape as part of the figure (e.g., triangle, square). The symmetry probes were included in large part simply to encourage the subjects actually to form images of the figures, whereas the shape probes were intended to assess a subject's ability to inspect the images looking for a part. For the two occurrences of the figure in each booklet, one had a "true" probe question and one had a "false" probe question; both types of probes occurred equally often when the probe was presented directly and when it was presented only after the figure had been reorganized. There were two symmetry questions for each booklet, one true and one false. In the reorganization condition, there were four true probe questions for each set of ten figures (eight in total from both booklets) which queried the presence of a particular geometric shape; one of the true probes named a part included in the initial description, one named a part included only in the revised description, one named a part included in both descriptions, and one named a part that had not been given in either description. The trial types were distributed this way because we wanted to encourage subjects to actually reorganize the figures and we wanted to vary the difficulty of "seeing" various parts in the images. In the direct probe condition half of the "true" probes named shapes that were specified in the description and half did not.

Procedure

All subjects were tested first on the booklet using the contiguous shape descriptions and then on the booklet using the overlapping forms descriptions. Prior to the experiment each subject was given a short definition of each elementary geometric shape that could be used to compose a form and a definition of the three types of symmetry that could be queried. Subjects were told that the first number presented on a trial corresponded to a page number in the booklet, and they should turn to that page, and then should read the description written on it. This description told them how to "see" the figure about to be presented. Upon hearing the next page number they were to turn to that page and to study the figure, seeing it as composed of the shapes noted in the accompanying description. When they heard the word "cover" they were to cover the figure with a small piece of cardboard. As soon as they heard the word "image," they were to form a mental image of the figure, keeping their eyes closed. As soon as the image was as clear and vivid as possible, they were to push the telegraph key under their dominant hand. The time taken to form the image was recorded by the experimenter. The subjects then either received an alternate description or a probe. On the reorganization trials the subjects were to reform their images immediately, now seeing it as composed of the new parts. As soon as the subjects were able to see the imaged figure in terms of the new parts, they were again to push the button under their dominant hands. This reorganization time was recorded. Following this, or instead of this, the subjects heard a probe question. Upon hearing the probe, the subjects were to search the image until they saw the geometric shape or axis of symmetry named, or until they could see that it was not present. The subject was to press the appropriate key, labeled "true" or "false," as soon as a decision was reached. All "true" responses were indicated by pushing the button under the subject's dominant hand, and "false" responses were indicated by pushing the other button. The judgments and response times were recorded by the experimenter. It was emphasized to the subjects that the probe trial was an "imagery detection task," and that we were interested in how long it took the subject to see the shape or axis of symmetry or to look and decide that it was not present.

The Models

Figure 16 presents our model of the image generation task. The first part of a figure is imaged, which serves as a skeleton. The PUT process then uses the FIND and PICTURE processes to locate where each subsequent part belongs and then to image it there.

Our measure here is the average increase in time to form images as each additional part is added (i.e., the slope of the image time over number of parts). This measure is clearly sensitive to differences in the speed of all three processes in the different conditions (i.e., with more or less complex stimuli.) Hence, our criterion leads us to weight all three processes.

Figure 17 presents our model of the image reorganization task. The image is first regenerated prior to being operated on. The FIND process is then used to locate a new unit (i.e., one mentioned in the new description). If it fails to find one before the image has faded, the PICTURE process is used to fill in the missing parts and the FIND process tries again.[6] Once a configuration of lines corresponding to a new part has been found (e.g., segments of adjacent squares that can be composed into a rectangle), the PARSE process is used. This procedure regenerates only selected segments, in this case just those segments that should be organized together. Thus, these segments are at the same "fade phase," and become organized together in acordance with the Gestalt law of common fate. Once a new unit is parsed, the image is regenerated and the entire procedure is repeated until the figure is completely reorganized.

Our measure in the reorganization task is the time to reorganize the image. Previous research has shown that variations in the complexity of an image affect inspection and generation times (see chapters 7 and 6 of Kosslyn, 1980). In this task variations in stimulus complexity will have effects on time because of the speed of the FIND, REGENERATE, and PICTURE processes, and hence all of these processes were weighted. Further, given that it should be more or less difficult to parse the different images, our rule for assigning weights also leads us to weight the PARSE process in this task.

Figure 18 presents our model of the image inspection process which leads to finding a part (this model is not complete for trials in which "false" properties were presented). The image is regenerated and the FIND process is used to search for the named part. If the image is complex or the part is composed of segments from different units in the image, the image may fade before the part is found. In this case the PICTURE process will be used to fill in the image again, and the inspection procedure will continue until the part is found. Our measure here is the time to find a part, and the stimuli vary in complexity and part-type. Given that the stimulus variations are known to affect inspection times, and times are a product of the speed of all of the components, all of them are weighted.

INSERT FIGURES 16, 17, and 18 ABOUT HERE

Results

Group Results

The data from all 50 subjects were considered in separate analyses of variance examining generation times, reorganization times, and probe times.

Generation Times: In the analysis of the initial generation times we examined the effects of contiguous vs. overlapping descriptions, number of units in the initial description, experimenter, and the sex of the subject. First, as is evident in Figure 19, the results replicated earlier findings on image generation times (see Kosslyn, 1980), with more time being required to generate images presumably composed of more units, $F(5, 170) = 32.77$, $p < .001$. This result was also evident in the generally longer times to generate contiguous units instead of overlapping units, $F(1, 34) = 58.85$, $p < .0001$. Thus, we have reason to believe that these data do in fact reflect the time to generate images. The difference in time taken for subjects to image figures using contiguous vs. overlapping descriptions was greater for female subjects, $F(1, 34) = 4.98$, $p < .05$, and for subjects tested by the female experimenter, $F(1, 34) = 9.47$, $p < .01$. These results were also apparent in a three-way interaction between type of description, experimenter, and sex of subject, $F(1, 34) = 5.55$, $p < .05$, with the greatest difference in generation times seen by female subjects tested by the female experimenter. No other effect or interaction was significant, $p > .1$ in all cases.

INSERT FIGURE 19 ABOUT HERE

Reorganization Times: In the analysis of the reorganization times we examined the effects of a contiguous vs. overlapping description, number of units in the second description, change in number of units from the original to the second description, the experimenter, and the sex of the subject. Subjects took longer to reorganize an image from the contiguous to the overlapping (3143 msec) description than vice versa (2102 msec), $F(1, 34) = 36.48$, $p < .0001$. This finding is evidence that subjects were in fact reorganizing their images, given that it confirms a prediction of the theory. In addition, although the reorganization times did vary with the number of units in the reorganized image, $F(3, 120) = 17.12$, $p < .0001$, these times largely reflected the difficulty of performing the reorganization. The figures with the smallest number of units (two) after reorganization were reorganized from contiguous to overlapping and took the greatest amount of time. The figures with the greatest number of units (five and eight) after reorganization were reorganized from overlapping to contiguous, and these reorganization times were shorter than the contiguous to overlapping reorganization times. However, when going from overlapping to contiguous organizations, more time was required to organize figures into eight units than into five units, which is as expected if each unit formed requires an increment of time. Those figures with an intermediate number of units (three) were reorganized both from overlapping to contiguous and from contiguous to overlapping, and the time to reorganize these images was in between the other two cases. Male subjects took longer to reorganize their images than did female subjects, $F(1, 34) = 10.23$, $p < .01$. Further, there was a greater difference in times for male subjects to reorganize their images from contiguous to overlapping descriptions than from overlapping to contiguous descriptions, $\underline{F}(1, 34) = 4.27$, $p < .05$.

Lastly, reorganization times were found to differ with changes in the number of units between the original to the reorganized description, $\underline{F}(4, 136) = 16.74$, $p < .0001$. Again, these differences varied with the direction of reorganization, with contiguous-to-overlapping reorganizations taking more time. For reorganization from overlapping to contiguous, times increased as the differences in number of units increased, with 2026, 2182, and 2326 msec for changes of one, three, and five units. This result supports the claim that subjects were parsing the larger overlapping units into smaller contiguous units one at a time, with about 75 msec being required for each parsing operation. In contrast, changes from contiguous to overlapping did not increase with the difference in the number of units, with 3301 and 3103 msec for changes of two and three units. This result makes sense because in this case the number of reparsing operations is not directly related to the number of units in the final organization.

Inspection Times: Probe times were examined in two separate analyses, first for cases in which subjects first had imaged the figure using only the initial description, second for cases in which subjects had to reorganize their images. In the first analysis we considered whether the shape mentioned in the probe had been included in the description or not. In the second analysis we compared the cases in which shapes were not mentioned in the descriptions at all, were mentioned in the originial description only, were mentioned in the second description only, or were mentioned in both descriptions. In both the non-reorganized and reorganized cases we examined the effects of image complexity, experimenter, and sex of the subject.

In the non-reorganized condition, there was no difference in probe times for trials on which the part was in the description and trials on which the part was not mentioned, $F(1, 46) = 2.21$, $p > .1$. However, probe times generally were faster for images described in terms of overlapping shapes (2160 msec) than contiguous shapes (2637 msec), $F(1, 46) = 15.08$, $p < .001$. This finding supports the inference that subjects were in fact inspecting their images, given that more complex images tend to be more degraded and more difficult to inspect (see Kosslyn, 1975). There was a tendency for subjects tested by the male experimenter to be faster, $F(1, 46) = 3.57$, $.05 < p < .1$; no other effects or interactions approached significance, $p > .1$ in all cases.

In the reorganized condition, subjects took longest to answer the probe question when it was a part not mentioned in either description, and took the least amount of time when the part had been used in both descriptions, $F(3, 138) = 10.79$, $p < .0001$. However, in contrast to the probes of non-reorganized images, once an image was reorganized it took more time to inspect it if it was composed of overlapping patterns (2993 msec) instead of contiguous ones (2486 msec), $F(1, 46) = 19.17$, $p < .001$. Recall that more time was required to reorganize the image from contiguous to overlapping than vice versa; apparently, images reorganized into overlapping shapes often were of poor quality and difficult to inspect. In addition, differences in the amount of time taken to answer probe questions in the different conditions (with the probe mentioned in the initial description, in the second description, in both descriptions, or in neither) depended on how a figure was described, in terms of contiguous or overlapping shapes, $F(3, 138) = 5.02$, $p < .01$. As is shown in Figure 20, the greatest difference in reaction times occurred when the probed shape was mentioned only in the reorganization description, with probe questions taking longer to answer when figures were imaged in terms of overlapping shapes (after reorganization) than when figures were imaged in terms of contiguous shapes.

These group results are difficult to interpret in detail, but are sufficient to serve present purposes: They do support our claims that times reflect image inspection time, and that a host of stimulus factors do affect processing (and hence lead us to weight the appropriate processes in the models).

INSERT FIGURE 20 ABOUT HERE

There was a mean of 10.8% errors overall. For the non-reorganized trials, mean errors were 6.7% when the probed part was in the description and 12.0% when it was not for images described in terms of contiguous shapes, and 1.0% and 14.0% for the two respective conditions for images described in terms of overlapping shapes. For the reorganized trials, mean error rates for probed parts not in any description, only in the initial description, only in the second description, and in both descriptions were 44.0%, 6.0%, 6.0%, and 2.0%, respectively, for figures initially described in terms of contiguous shapes, and 44.0%, 8.0%, 0.0%, and 0.0% for figures initially described in terms of overlapping shapes. Errors for trials on which the probed part was not in the figure were 10.6% and 9.6% for the contiguous and overlapping images, respectively. Thus, errors did not increase as times decreased, belying any speed-accuracy tradeoffs.

Individual Difference Measures

We created four measures that indexed important aspects of the image generation, reorganization, and inspection processes. Each subject received a z score on each of these measures, with faster times or slopes receiving higher z scores. The descriptive statistics for each measure are as follows:

The first measure was the slope of the best fitting function for the times to generate forms with varying numbers of units. This measure indexed the speed with which additional parts could be placed on an image. Three subjects produced negative slopes, suggesting that they were not doing the task. If the task had been easy for them, we reasoned, they would have been likely to do it. Thus, we assigned these three people z scores .01 below the lowest obtained from the other subjects. Excluding these data, the mean slope was 191 msec per unit, with a standard error of the mean of 40 msec. The minimum slope was 2 msec per unit and the maximum was 1715 msec per unit.

The second measure was the speed with which subjects could reorganize their images. The mean time was 2622 msec, with a standard error of the mean of 178 msec. The maximum mean time was 6988 msec, and the minimum was 830 msec.

The next measure was the mean time to inspect a non-reorganized image for a part, and the final measure was the mean time to inspect a reorganized image for a part. Only times from correct "true" responses (where the part was presumably found in the image) were used in calculating these means. For the non-reorganized trials the mean response time was 2403 msec, with a standard error of the mean of 134 msec. The maximum mean response time was 5482 msec, and the minimum was 1062 msec. For the reorganized trials the mean response time was 2946 msec, with a standard eror of the mean of 225 msec. The maximum response time was 9025 msec and the minimum was 1017 msec. The two probe times were considered separately because different stimulus factors contributed to the times.

Discussion

The results from all three sub-tasks provide evidence that the subjects actually used imagery as requested. First, more time was required for subjects to form images of forms containing more parts, as is commonly found in image generation tasks (see Kosslyn, 1980, Chapter 6). However, the slopes here were about 50% steeper than those usually found in the same task with college students. Second, the results from the reorganization trials confirmed a prediction of the Kosslyn and Shwartz theory of imagery, with more time being required to reorganize contiguous forms into overlapping ones than vice versa. Further, only when overlapping forms were reorganized into contiguous ones did the difference in the number of units between the two matter--which is exactly as expected if each unit was redivided and each parsing operation required an increment of time. No such relation should hold for the reorganization of contiguous to overlapping forms, since the primary determinant of reorganization time here will be the number of operations required to locate segments on different units. Finally, the results from the probe task replicated the usual finding that more time is required to "see" parts of more complex images (see Kosslyn, 1980, Chapter 7). This was not true when subjects first reorganized their images, however, which may reflect the fact that more complex reorganized images took less time to reorganize and hence may have been less faded and more easily inspected.

One somewhat intriguing finding here was the existence of sex differences in the generation and reorganization tasks. Given that there were no sex differences in the inspection task, it seems unlikely that males and females differ in their use of the FIND, PICTURE or REGENERATE processes. We have no good account of these findings, but note that the PARSE process is required in "embedded figures" tasks, in which a figure is hidden in a network of interconnecting lines. Performance on these sorts of tests is related to "field dependence," a relative inability to disassociate a figure from its context, and women tend to be more field dependent than men (e.g., Witkin, 1964). Thus, it is interesting that <u>men</u> appear to have been poorer at using the PARSE process, suggesting that it is not at the heart of sex differences in field dependence.

Written Tests

At the conclusion of the final session each subject was given five written tests and one questionnaire; one of the tests and the questionnaire have been taken to assess imagery ability, and thus these scores were analyzed along with the measures collected in our other tasks. The remaining tests have not been taken to reflect imagery ability, and thus these scores were analyzed later with an eye toward discovering whether individual differences in imagery reflect differences in more general cognitive ability.

Form Board Test

The first test administered was the Form Board Test (VZ-1). This test was taken from the "Kit of factor referenced cognitive tests," prepared by Ekstrom, French, and Harman (1976). The test was administered according to the directions included in the kit, with subjects always reading the instructions and the experimenter allowing them only the allotted time. This test requires a person to decide which of five two-dimensional shapes could be fitted together into a pattern that matched a standard figure. There were four different figures to be constructed, with six sets of five shapes each, producing a total of 24 problems (because of time constraints we used only part 1 of the test, deleting the second 24 items). Subjects examined each set of five shapes and marked with a plus the shapes that would fit together to form the standard figure and marked with a minus the shapes that would be omitted in forming the standard. Subjects were allowed a total of eight minutes to finish the test.

The Model

Unlike our own tasks, which were designed to be relatively simple, there were multiple possible models for performance of this test. We chose the simplest one, which is presented in Figure 21. The LOAD process first encodes a part, and then the image is compared to the shape of the standard figure. If there is no possible way it could fit (e.g., is too tall), a response is made for that form and a new form is encoded. When there is a possibility the form could fit, the TRANSLATE and ROTATE processes are used to explore possible fits, until the FIND process determines whether or not it will fit. When a part fits, a positive response is made. When an item fits, an image of it must be maintained in the proper location in the standard figure, because the composite of the parts imaginally inserted into the form constrains the "space" available to fit other forms. Thus, the REGENERATE process is used to maintain the image prior to using LOAD to encode a new part. This procedure continues until the form is filled with imaged parts.

Our method for determining weighted components depends on factors that make some items easier than others. Given that the task is timed, and hence speed is important, these factors are the similarity of forms to the standard, the amount of rotation and translation necessary, and the number and complexity of the forms that must be maintained in the image over time. We had evidence that these stimulus factors affect performance in similar tasks, and our theory localized these affects to all but the LOAD process (we have no evidence that the differences in part shape per se were more or less difficult to load into an image). Thus, all but the LOAD process are weighted.

INSERT FIGURE 21 ABOUT HERE

Results

Each subject's score was the total number marked correctly minus the total number marked incorrectly. This measure allows "partial credit" for combinations that were nearly correct. For purposes of later analyses, the scores were converted to z scores.

The mean score on this test was 34.5 (of 120 possible), with a standard error of the mean of 2.74. The maximum score was 80.0, the minimum was -3.0. Thus, the test was very difficult. Many subjects attempted as few as three of the 24 sets of five, and did not always get these problems correct. Higher scores received higher *z* scores.

VVIQ Questionnaire

The subjects were asked to complete the Vividness of Visual Imagery Questionaire (VVIQ; see Marks, 1977). This questionnaire consists of a set of descriptions of scenes which the subject is to image: For example, "The sun is rising above the horizon into a hazy sky" and "The color and shape of a lake." The subject is to rate the vividness of each image, using a standard five-point rating scale, with "1" indicating "perfectly clear and as vivid as normal vision" and "5" indicating "no image at all, you only 'know' that you are thinking of the object." This questionnaire is a refinement of the visual scales of the original Betts Questionnaire on Mental Imagery (see Richardson, 1969), and has been shown to predict individual differences in the performance of a number of imagery tasks (see Finke, 1980; Marks, 1977). Although there was no set time limit, subjects were told that it should take them only five to ten minutes to complete the 32 items (16 items, once imaged with eyes open and once imaged with eyes closed). This questionnaire purportedly assesses the vividness of a person's images.

The Model

Figure 22 presents our model for the VVIQ task. The subject must use the PUT, FIND and PICTURE processes to form a composite image as described. Once formed, the image is regenerated to be as sharp as possible and the RESOLUTION process is used to assess its vividness.

Determining which components should be weighted in this task was extremely difficult; we had no data about which factors systematically affect this performance measure in this task, and hence could not reliably determine which processes were mitigating the effects of stimulus factors. Logically, however, it seemed clear that differences in the RESOLUTION process would affect differences in apparent vividness, as would differences in the efficiency of the PICTURE process which produces an image in the visual buffer. Thus, it seemed most conservative to weight only these two processes.

INSERT FIGURE 22 ABOUT HERE

Results

The subject's score was the mean of the ratings for the items. Z scores were computed for the eyes-open and eyes-closed conditions averaged together. The mean was 70.4 with a standard error of the mean of 3.4. The maximum score was 147.0, the minimum was 34.0.

Non-imagery Tests

Four additional tests were administered from the Ekstrom et al. kit. None of these tests seemed to require visual imagery; thus, if our components are imagery-specific and imagery ability is independent of other cognitive abilities, then scores on these tests should not be systematically related to scores on the previous tasks and tests. The Auditory Number Span Test (MS-1) assessed how many digits a person could hold in short-term memory; the Extended Range Vocabulary Test (V-3) assessed how well subjects could identify synonyms; the Different Uses Test (XU-4) assessed how many uses subjects could associate quickly to common items (soap, barrel, sock and snap); and the Nonsense Syllogisms Test (RL-1) assessed how well subjects could determine if specific conclusions were implied by nonsensical premises in simple syllogisms (e.g., "All trees are fish;" "All fish are horses;" Therefore, "All trees are horses").

III. COMPONENTIAL ANALYSES

We now are in a position to address the questions that motivated us to do this study, namely is imagery ability general and undifferentiated or is it a consequence of a set of relatively independent subabilities? And if the latter, do these subabilities correspond to those posited in the Kosslyn and Shwartz theory? We began to address these questions by correlating the z scores assigned to each person across the different tasks, producing the matrix presented in Table 1. If imagery ability were general and undifferentiated (i.e., people were either "good" or "poor" imagers in general, as is often assumed in the literature), then these correlations should have been uniformly high. This clearly was not the case. Further, as is evident in Table 2, the split-half reliabilities of the individual tasks are reasonably high (with a mean of .78), and are much higher than most of the correlations. Thus, the low correlations evident in Table 1 are not due to error of measurement or exceptionally noisy data.[7] In addition, the correlations are not simply uniformly low, as would be expected if each task recruited an entirely independent ability. Rather, there was a a range between -.44 and .79, as would be expected if the tasks shared greater or lesser numbers of the same underlying components. Presumably, the more shared components, the higher the correlation. Negative correlations arise when non-shared components are negatively correlated (as could occur, for example, between the FIND and REGENERATE processes, with people who have more sensitive FIND processes not needing to develop good image retention abilities). According to our theory, such negative correlations among the efficacy of individual component processes should be increasingly less likely to show up as negative correlations among measures of task performance when the tasks have more components in common, because the influence of the common components should lead to a positive correlation, overshadowing the influence of the remaining non-shared components.

INSERT TABLES 1 AND 2 ABOUT HERE

If our tasks were in fact performed using a relatively small set of distinct components, and these component structures and processes are specific to image representation and use, then we would not expect correlations between our measures and scores on non-imagery tests. As is evident in Table 3, this was by and large true. First, for nine of the thirteen imagery tasks there were no significant correlations between _z_ scores on our performance measures and _z_ scores on the non-imagery tests, and for two more of our tasks there was a relatively low correlation with one of the non-imagery tests. For these 11 tasks, then, there were about as many significant correlations as would be expected due to chance alone. In contrast, the line drawing scores were clearly related to the non-imagery test scores, with all four of them being significantly correlated. Similarly, the Form Board scores were significantly correlated with two of the non-imagery test scores. Both the line drawing task and the Form Board test purportedly critically require using the REGENERATE process to maintain information over time. This ability may be enhanced if one is good at "chunking" the image, organizing it into fewer units. Such chunking ability may reflect a more general "intelligence" factor that is also tapped by the non-imagery tests (but not, please note, by the other imagery tasks). This notion is consistent with the fact that there were significant correlations between every possible pair of the non-imagery test scores but one. In any event, it is clear that in general our imagery tasks are not simply tapping a general undifferentiated imagery ability or a general cognitive ability.

INSERT TABLE 3 ABOUT HERE

Given the evidence that the imagery tasks recruited distinct processing components, it makes sense to ask whether the Kosslyn and Shwartz theory adequately characterizes these components. Table 4 presents a summary of the processing components assumed to underlie performance in each task: weighted components are capitalized for each task. We began by estimating the similarity in the processing of each pair of tasks using a very simple procedure: For each pair, we counted how many different components were used; this was treated as a denominator. We then counted how many components were purportedly shared in the two tasks. Further, if a component was shared and was weighted in both tasks, we added two more points to the numerator; this number was chosen because on the average tasks had about two non-weighted components, and we considered one weighted component to be at least as important as the non-weighted components in a task. In addition, if a component was weighted and used in only one task, we added two points to the denominator. The total numerator score was divided by the total denominator, producing the values presented in Table 5.[8]

INSERT TABLES 4 AND 5 ABOUT HERE

The first measure of the adequacy of the Kosslyn and Shwartz theory was the simple correlation between the predicted and observed similarities (i.e., Tables 1 and 5). This correlation was $r = .56$, $p < .0001$. This correlation is highly significant, and must be evaluated in light of the average split-half reliability of $r = .78$, which provides one measure of the systematic variance in the data. In a second analysis, we simply eliminated the Form Board written test and the VVIQ questionnaire. The Form Board Test is quite complex and clearly can be performed using numerous strategies, only one of which (the simplest, in our judgment) was considered here. For example, one could encode each part into long-term memory and then use the PUT, PICTURE and FIND processes to try to fit the parts into the standard figure, instead of using the LOAD process followed by various transformations, as we have assumed. The VVIQ test requires a judgment rating that is difficult to relate to the underlying processing components, which kept us particularly uncertain about our weighting assignments. Thus, we decided also to examine only those tasks for which we had confidence in our models. The correlation between the predicted and observed similarities was now $r = .67$, $p < .0001$.

We next examined our claim that the efficiency of the weighted components was more important for determining performance than was the efficiency of non-weighted components. Now we simply compared the weighted components in each task, adding 1 to the denominator for each additional weighted component used in the two tasks (i.e., used in either model) and adding one to the numerator for each additional shared weighted component (i.e., one used in both models). The correlation between this measure of similarity and the observed correlations in task performance was $r = .61$ when all tasks were considered, and $r = .78$ when the Form Board and VVIQ were eliminated. Clearly, variation in the weighted components is primarily at the root of the observed individual differences in task performance. Although not perfect, the Kosslyn and Shwartz theory characterizes the underlying processing components at least reasonably well.

Probably the best way to evaluate our theory is to contrast it with a plausible alternative theory. Unfortunately, there is no alternative theory for image processing, and thus we had to make up our own. We began by considering image <u>vividness</u> and <u>control</u>, two factors often assumed to underlie individual differences in imagery ability (see Richardson, 1969). In addition, a person's <u>speed</u> and <u>memory</u> ability seemed intuitively plausible as factors that affec results in our tasks. Two of us familiar with the tasks categorized each task intuitively in terms of whether or not each factor seemed relevant; this categorization was done before the actual correlation matrix was obtained. These categorizations did not rest on any of the assumptions of the theory, such as those concerning the role of the FIND process in image generation.

The alternative "commonsense theory" (CT) was used to generate a new set of predicted similarities among all pairs of tasks. First, we simply computed a measure of similarity for each pair of tasks as we did with our models. The numerator was the number of shared components (i.e., shared positive values), and the denominator was the total number of different components (i.e., positive values) involved either task. The proportion of common values correlated with the observed correlations only r = .30. But this seemed hardly fair, given that the CT theory was so much more simple than ours; thus, we proceeded to make more subtle comparisons using regression analysis (cf. Sternberg, 1977).

In our first regression analysis we created four independent variables. For each pair of tasks, the vividness, control, speed, and memory requirement scores presented on the right side of Table 4 were compared. If the values were the same (both + or both -), a 2 was assigned to that factor; if they were different, a 1 was assigned. The observed correlations were regressed against these values. The results were as follows: First, the simple correlations between each factor and the observed correlations among the tasks were: vividness, r = .18; control, r = .04; speed, r = .21; memory, r = .01. The multiple correlation was r = .29.

The comparable analysis was done using our model. Now a 4 was assigned for a component if it was weighted and shared by two tasks, a 3 if it either was present in both or not present in both, a 2 if it was in one but not the other, and a 1 if it was in one and weighted but not in the other. The multiple r was now .55. The similarity of the multiple correlations and the correlations observed in our initial analysis (comparing predicted task similarity and observed correlation) is, of course, exactly as one would expect and gives us confidence in this regression procedure.

There are two problems with comparing the regression analyses evaluating the two theories: First, our theory has many more parameters. However, this theory was not invented just to explain the present data; rather, these components were necessary to provide accounts for a wide range of data having nothing to do with individual differences (see Kosslyn, 1980). Nevertheless, it was of interest to concoct a slimmed down version of our theory, containing only the four processes that were weighted most often, the FIND, PICTURE, REGENERATE and RESOLUTION components, which reflected all but the transformation subabilities. The second problem in comparing the two regression analyses lies in the fact that only binary values were used for CT theory whereas a range of weightings (1, 2, 3 or 4 values) were used for our theory. We therefore reanalyzed the data, this time examining only the four processes noted above, and indicating with a 2 when two tasks shared a component <u>and</u> it was weighted in both tasks and a 1 otherwise. The multiple $\underline{r}$ between the components and observed correlations was .60. Even this skimmed-down version of our theory was superior to the intuitive common sense one, based on easily-observable surface pictures of the tasks (such as whether speed or memory ability was important).

All of the foregoing correlations are improved if we ignore the Form Board and VVIQ tests. The commonsense theory now produces an $\underline{r}$ of .48 with the observed correlations, compared to an $\underline{r}$ of .70 for our theory (and $\underline{r}$ = .82 for when we considered only the weighted relations among the four "basic" components). Performance on these two tests apparently reflects complex processing that is not being very well characterized by the models, and/or the tasks are measuring substantially different things in different people.

INSERT TABLE 6 ABOUT HERE

Our theory provides much better fits to the data than the commonsense theory, but largely because of the relations among weighted components. This observation is driven home by the results of another regression analysis of the data, where non-weighted components are simply ignored. Now a 2 is assigned if the tasks agree in including or not including a weighted component, and a 1 is assigned if one task includes a weighted component and the other does not. The multiple $\underline{r}$ here is .55 (.77 when VVIQ and Form Board are excluded). Given the importance of the weighting scheme, then, we decided to conduct an independent test of how good our weighting procedure was. Table 6 presents a summary of a regression analysis using no weights at all. Here, a 2 was assigned if two tasks both included or excluded a given component and a 1 was assigned if only one of the tasks included a component. Hence, the weightings were ignored altogether in this analysis. The Beta weights in Table 6 represent the relative importance of each component in accounting for the data, and those differences in relative importance ought to be reflected by the theoretical weightings we have assigned. That is, according to our theory, the reason some components have higher Beta weights than others is that these components are more often weighted when they occur, and hence when they occur they account for a greater amount of the similarities among task performance. To examine this conjecture we performed a simple test: For each component, we examined how many times it occured in the various tasks (i.e., in Table 4), and how many of those times it was weighted. We then computed a proportion, the number of times weighted relative to the number of times it occured, for each component that occured more than once (all but the PAN and PARSE process). These proportions were then simply correlated with the corresponding Beta weights. This correlation was $\underline{r}$ = .90. When the Form Board and VVIQ were not included, this correlation was $\underline{r}$ = .75--the drop in correlation in part reflecting the decrease in number of components (the TRANSLATE process now occured only once, and hence was not included) and the relatively higher contribution of the disparity between the predicted and observed contribution of the PICTURE process (evident in Table 6). It is not clear, of course, if we should have weighted this process in additional tasks or if we presently have weighted it in some tasks where it should not have been weighted. In any event, our weighting rule did a remarkably good job in assigning weights to the different components in the different tasks.

In short, then, not only does our theory do a creditable job in accounting for the patterns of variation in performance of the tasks, but it does better than a commonsense alternative theory.

<u>Cluster Analysis and Factor Analysis</u>

The correlation matrix presented in Table 1 was submitted to a hierarchical cluster analysis (AGCLUS, written by D. Oliver). In this analysis, the most similar (highly correlated, in our case) items are grouped into an initial cluster and other items are added until the maximally-similar items are included in a cluster, with the aim of also maximizing the inter-item similarity within the remaining clusters (see Johnson, 1966). The inter-cluster similarities, which are used to group the clusters hierarchically, are computed on the basis of the average similarity between items in a cluster (including the diagonal entries, which has the effect of biasing the solution toward clusters with small numbers of items.)

INSERT FIGURE 23 ABOUT HERE

Figure 23 presents the results of performing the hierarchical clustering analysis on the data. The shorter the lines connecting two measures into a cluster, the more tightly they are clustered. The first thing to notice about Figure 23 is that clearly demarcated clusters were produced, as expected if a relatively small set of components are used in image processing and tasks differ in the number of common components recruited. The second thing to notice is that the clusters are easily interpreted in terms of the components posited by the Kosslyn and Shwartz theory. The top cluster contains all tasks in which the FIND and PICTURE processes are weighted. It also includes the rotation slopes and VVIQ scores. This cluster, then, seems to reflect the efficiency of the FIND process in conjunction with transformation processes--either ones that produce images in the visual buffer on the basis of stored information or ones that manipulate the image in some way. The fact that the FIND process was not weighted in the VVIQ is probably best regarded as additional evidence of our failure to characterize adequately the model for the VVIQ. The next cluster, which is joined with the first under a superordinate node, contains all three of the tasks in which the REGENERATE process was required to maintain an image over a relatively long period of time. Thus, this cluster seems to indicate that the capacity (not speed) of the REGENERATE process was an important factor in task performance. Note also that the FIND process is used in each task, explaining why this cluster is joined with the first. Finally, the third major cluster seems to reflect structural properties of the visual buffer. The acuity, extent, and oblique tasks were designed to reflect bottlenecks imposed by the sensitivity of the RESOLUTION process and the grain of the visual buffer. This limitation on vividness has a different source, evidently, from that assessed by the VVIQ task (which may reflect variations in the RESOLUTION process, as opposed to the grain of the medium). The cluster including these three measures is included in a larger cluster which also contains the described scenes measure (the correlation between time and the distance that should have been scanned across). At first glance, this organization was a puzzle. One possible reason for this clustering, however, makes perfect sense: Most of the described scenes extended over around five inches (at arm's length subtending about 16 degrees of visual angle), which was greater than the extent of the visual buffer estimated by many of our subjects (recall that the minimum estimate was 1.2 inches, subtending around four degrees at arm's length). Thus, some subjects may have had difficulty in maintaining the correct distances in the described scenes because they overflowed the available extent of the visual buffer. In order to image the scene, then, they may have had to distort the distances among the objects, and hence produced low correlations between scan time and expected distance. In short, all four measures in this last cluster may reflect variations in properties of the visual buffer itself.

INSERT TABLES 7 AND 8 ABOUT HERE

Finally, we performed a factor analysis on this data, specifying that we wanted oblique axes rotated (direct quartimin, used in the BMDP package) into the best fit. This was the most conservative approach because we could not be certain that the efficiency of the various components was in fact independent. The first results of interest are presented in Table 7, which shows that the factors were not highly intercorrelated. The second results of interest are presented in Table 8, which presents the factor pattern matrix (with loadings reflecting the importance of each factor for each measure). The easiest way to interpret this matrix is to consider only values over some absolute magnitude, .2500 being the most common criterion. If we do this, we see that the two image reorganization times, image inspection times, described scenes, generation slopes and VVIQ all are weighted on the first factor. In every case the PICTURE process occurs in our model for the task, and this process is weighted in every case but the described scenes--which is the only measure that has a negative factor loading here. Note that this first factor is <u>not</u> simply speed of response. If it were, the rotation slopes, line drawing times and Form Board measures should have been weighted but were not, and the VVIQ measure should not have been included, but it was. Similarly, the four measures that load heavily on the second factor, acuity, extent, the oblique effect and the VVIQ, all include a weighted RESOLUTION process in their models. The VVIQ actually loads negatively on this factor, however, which is consistent with our results from the cluster analysis; the VVIQ apparently is not simply a measure of "image vividness," as has been assumed. This second factor clearly represents the importance of the RESOLUTION process. The third factor is more difficult to interpret, although the high loadings for the line drawing scores and Form Board suggest that it reflects the effectiveness of the REGENERATE process in maintaining our image over time. The importance of this process for the speed of making a later decision seems to be reflected by the fourth factor, as witnessed by the very high loading for the line drawing judgment times. The fifth factor is uninterpretable.

The first factor, then, corresponds roughly to the first cluster of our AGCLUS analysis; note that the three measures weighted most highly were the first three factors entered into the cluster; and the last factors entered (which were the least tightly clustered) have relatively low weightings on this factor. The second factor corresponds fairly closely to our third major cluster, both apparently reflecting the contribution of the RESOLUTION process and the grain of the visual buffer. The VVIQ clearly is a more complex measure than it is usually made out to be, with the RESOLUTION process not contributing solely to this measure. Finally, the third factor corresponds roughly to the second cluster, both apparently reflecting how much information can be maintained in an image over time.

The distinction between the first two factors provides additional support for our theory as opposed to something like the commonsense theory discussed earlier. Speed per se does not seem to be an important factor, nor does a commonsense idea of "image vividness" underlie the results. In general, the results of the factor analysis are consistent with the previous analyses. The one major puzzle is why the effectivenesss of the PICTURE process seemed so important here, yet did not seem very important in our regression analyses.

IV. CONCLUSIONS

The results of this study support two broad assertions: First, imagery ability is not general or undifferentiated. Rather, a number of relatively independent subabilities can be drawn upon when one uses visual mental imagery. Second, the Kosslyn and Shwartz theory lends insight into the underlying structure of imagery. This theory predicts the observed correlations among task performance reasonably well, and provides a straightforward way of interpreting the results of our cluster analysis.

The results of this study also underline another conclusion: The population at large is considerably different from the population of university sophomores usually tested in psychological research. We were repeatedly surprised at how poorly our subjects performed our tasks. One could argue that this is in part a motivational factor, that these people were not "good subjects." But if so, some surely would have been more motivated than others, and we would have expected results indicating simply that some people are generally better than others. This was not the case. One could also argue that subjects were motivated to perform only some of the tasks, with different subjects being motivated more or less for different tasks. But why would motivation differ so systematically across tasks? If the subject neglected those tasks that were especially difficult for him or her, then the fact that subjects did poorly on a task still serves to provide a measure of underlying ability.

Another result of the study was that practice improved performance on most of the tasks (see Table 2). This result raises two questions: First, would these improvements be retained over time? And second, would they generalize to new stimulus materials? For example, would training on image maintenance improve general performance in the future, and would it improve with new materials? Attempting to train people on the various imagery tasks is worthwhile not only from a practical point of view, but also from a theoretical point of view: According to the Kosslyn and Shwartz theory, certain factors, such as the grain of the visual buffer, ought not to be capable of being improved with training (these components are not "cognitively penetrable," to use the term of Pylyshyn, 1981, and hence are posited to be directly embodied in neural tissue; see Kosslyn, Pinker, Smith & Shwartz, 1979). Other factors, such as the accuracy and efficacy of the FIND, PUT, and RESOLUTION procedures, should improve with practice (see Kosslyn, et al., 1979).

The next stage in this research is to develop direct measures of the efficacy of the individual components themselves, and then to construct "imagery strengths profiles" for people, indicating their relative strengths and weaknesses. We should then be able to predict performance in tasks in which imagery is used spontaneously: Hopefully, people will do well in imagery tasks that make use of their particular information-processing strengths, and will do poorly in tasks that require use of components that are not very efficacious for that person. If this is true, we may eventually be in a position to begin to instruct people about how they, as individuals, should approach particular tasks. And this surely would be an integral part of a truly deep and comprehensive educational program.

Footnotes

Requests for reprints should be sent to Stephen M. Kosslyn, Department of Psychology, The Johns Hopkins University, Baltimore MD 21218. This work was supported by ONR contracts N00014-79-C-0982 and N00014-82-C-0166. The authors wish to thank Gail Pendleton for editorial and secretarial assistance, and Earl Hunt for comments on an earlier draft.

1. See Shwartz (1979) for evidence that at least one image transformation, mental rotation, is performed on the image itself (as opposed to the underlying representations in long-term memory).

2. In Kosslyn (1980) two executive processes, IMAGE and LOOKFOR, are also discussed. It now seems that much of the control they assumed can be achieved locally (by specifying the input/output characteristics of the individual processes more precisely) or by a more general executive (not specific to imagery). Given this possibility, it makes more sense to take seriously only the specific processing components described above.

3. The ethnic backgrounds included Anglo, Black, Irish, Italian, Hispanic, Jewish, and Portuguese, the predominant groups in the Cambridge/Boston area. Their educational level ranged from one man who completed the 11th grade of high school to a Ph.D. candidate in Art History. The majority of subjects had completed a bachelor's degree, and several had Master's degrees. According to the subjects, their occupations included cook, story-teller, English tutor, clerk, TV station owner, dancer, mother, geriatrician, artist, EEG technician, writer, engineer, stockboy, secretary, and student (only 12 full-time students were tested, five of whom attended Harvard/Radcliffe).

4. The functional vividness or resolution of an image will in fact reflect two factors: the grain of the visual buffer and the sensitivity of the RESOLUTION process. When speaking of properties of the process we are in fact speaking of its properties in the context of properties of the visual buffer.

5. We do not have a theory for the oblique effect per se, but merely assume it reflects a peculiarity in the structure of the visual buffer and/or the operation of the RESOLUTION process.

6. The REGENERATE process can only maintain images in the visual buffer; once they have faded, it cannot bring them back from long-term memory. In the computer simulation model the parameters used to form an image of each part are stored temporarily, associated with the long-term memory encoding of the part, allowing us to avoid using the PUT and FIND processes to generate an image a second time immediately after it has faded. We assume that this is also true of people, allowing them to use the PICTURE process to re-generate an image from long-term memory. However, the PICTURE process can be used to re-generate an image only if the requisite encodings are in long-term memory. In the line drawings task subjects presumably did not have time to encode most of the directions in long-term memory, and hence had to rely on the REGENERATE process.

7. The split-half reliabilities were computed by calculating separate scores for the first and second halves of the trials for each task, and then correlating these scores across subjects. In the mental rotation and image inspection tasks there were slightly more observations in one half than the other because an odd number of trials was used in each condition.

8. The measure used in the rotation task was the speed of rotation, which is completely independent of the efficiency of the REGENERATE process used to refresh the image after it is rotated, prior to being classifed (this component would contribute to the intercept of the function, not the slope). Thus, in all analyses reported here we have deleted this component from this one model, which clearly could not have contributed to the performance measure we used. (Similarly, given our various measures we omit components involved in initial processing prior to forming an image that later is used--for example, image generation components are not considered when we measure the time

to inspect an image after it has been formed.)

References

Appelle, S. Perception and discrimination as a function of stimulus orientation: the "oblique effect" in man and animals. Psychological Bulletin, 1972, 78, 266-273.

Beech, J. R. Imaginal vs. perceptual scanning of a visual representation. Perception and Motor Skills, 1980, 50, 367-370.

Betts, G. H. The Distribution and Functions of Mental Imagery. New York: Columbia University Press, 1909.

Bower, G. H. Mental imagery and associative learning. In L. Gregg, (Ed.), Cognition in Learning and Memory. New York: Wiley, 1972.

Cooper, L. A. Mental rotation of random two-dimensional shapes. Cognitive Psychology, 1975, 7, 20-43.

Cooper, L. A., and R. N. Shepard. Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), Visual Information Processing. New York: Academic Press, 1973.

Ekstrom, R. B., French, J. W., and H. H. Harman. Manual for kit of factor-referenced tests. Princeton, N. J.: Educational Testing Service, 1976.

Farah, M. J., and S. M. Kosslyn. Structure and strategy in image generation. Cognitive Science, 1981, 4,

Fechner, G. T. Elements of Psychophysics. New York: Holt, Rinehart, and Winston, 1966 (originally published 1860).

Finke, R. A. Levels of equivalence in imagery and perception. Psychological Review, 1980, 86, 113-132.

Finke, R. A., and S. M. Kosslyn. Mental imagery acuity in the peripheral visual field. Journal of Experimental Psychology: Human Perception and Performance, 1980, 6, 126-139.

Galton, F. Inquiries into Human Faculty and its Development. London: MacMillan, 1883.

Gordon, R. A. An investigation into some of the factors that favour the formation of stereotyped images. British Journal of Psychology, 1949, 39, 156-167.

Johnson, S. C. Hierarchical clustering schemes. Psychometrika, 1967, 32, 241-254.

Kosslyn, S. M. Scanning visual images: some structural implications. Perception and Psychophysics, 1973, 14, 90-94.

Kosslyn, S. M. Information representation in visual images. Cognitive Psychology, 1975, 7, 341-370.

Kosslyn, S. M. Measuring the visual angle of the mind's eye. Cognitive Psychology, 1978, 10, 356-389.

Kosslyn, S. M. Image and Mind. Cambridge, MA: Harvard University Press, 1980.

Kosslyn, S. M. The medium and the message in mental imagery: a theory. Psychological Review, 1981, 88, 46-66.

Kosslyn, S. M., Ball, T. M., and B. J. Reiser. Visual images preserve metric spatial information: evidence from studies of image scanning. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 47-60.

Kosslyn, S. M., Reiser, B. J., Farah, M. J., and S. L. Fliegel. Generating visual images. Submitted for publication.

Marks, D. F. Visual imagery differences and eye movements in the recall of pictures. Perception and Psychophysics, 1973, 14, 407-412.

Marks, D. F. Imagery a consciousness: a theoretical review from an individual differences perspective. Journal of Mental Imagery, 1977, 1, 275-290.

Paivio, A., Yuille, J. C., and S. Madigan. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph Supplement 76, 1968.

Pennington, N., and S. M. Kosslyn. The "oblique effect" in mental imagery. Submitted for publication.

Pinker, S. Mental imagery and the third dimension. Journal of Experimental Psychology: General, 1980, 109, 354-371.

Pinker, S., and S. M. Kosslyn. The representation and manipulation of three-dimensional space in mental images. Journal of Mental Imagery, 1978, 2, 69-84.

Pylyshyn, Z. W. The imagery debate: analogue media versus tacit knowledge. Psychological Review, 1981, 87, 16-45.

Richardson, A. Mental Imagery. New York: Springer Publishing Co., 1969.

Shepard, R. N. The mental image. American Psychologist, 1978, 33, 351-377.

Sternberg, R. J. Intelligence, Information Processing, and Analogical Reasoning: the Componential Analysis of Human Abilities. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1977.

Shwartz, S. P. Studies of mental image rotation: implications for a computer simulation of visual imagery. Ph.D. Thesis, The Johns Hopkins University, 1979.

Weber, R. J., and M. Bach. Visual and speech imagery. British Journal of

Psychology, 1969, 60, 199-202.

White, K., Sheehan, P. W., and R. Ashton. Imagery assessment: a survey of self-report measures. Journal of Mental Imagery, 1977, 1, 145-170.

Witkin, H. A. Origins of cognitive style. In C. Sheerer (Ed.), Cognition: Theory, Research, Promise. New York: Harper & Row, 1964.

Figure Captions

Figure 1. Model for the acuity task. In this and all later models, capitalized letters represent weighted components whereas lower case letters represent non-weighted components.

Figure 2. Model for the oblique effect task.

Figure 3. Group results for the acuity task, with mean distance at the point of blur considered for different stimuli and different trials.

Figure 4. Model for the extent task.

Figure 5. Group results for the extent task, with mean distance between the stimuli compared for the different trials.

Figure 6. Model for the rotation task.

Figure 7. Group results for the rotation task, with decision time considered when stimuli were presented at different degrees of tilt.

Figure 8. Group results for the rotation task, considered for each stimulus and decision type (averaged over amount of rotation).

Figure 9. Model for the line drawing task. (This model reflects only factors that should affect drawing performance.)

Figure 10. Model for the line drawing judgment task. (This model assumes that the image has been formed prior to this task.)

Figure 11. Group results from the line drawing task, with percentage of lines drawn correctly considered for each trial.

Figure 12. Group results for the line drawing judgment task, considering number of line segments in the image and trial.

Figure 13. Model for the described scenes task.

Figure 14. Group results from the described scenes task, with distance of scan considered for the two directions.

Figure 15. Stimuli used in the image generation, reorganization, and

inspection tasks. Each stimulus can be described in terms of overlapping, relatively few parts or contiguous, relatively many parts.

Figure 16. Model of the image generation task.

Figure 17. Model of the image reorganization task.

Figure 18. Model of the image inspection task (in cases in which the sought part is in fact found).

Figure 19. Group results for the image generation task, with image generation time when different numbers of parts were in the imaged figure.

Figure 20. Group results for the image inspection task after the image was reorganized, with type of image organization considered for each description condition.

Figure 21. Model for the Form Board Test.

Figure 22. Model for the VVIQ questionnaire task.

Figure 23. Results of the AGCLUS analysis.

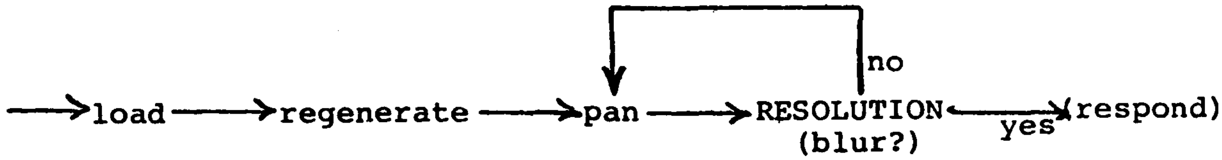


Figure 1. Model for acuity task.

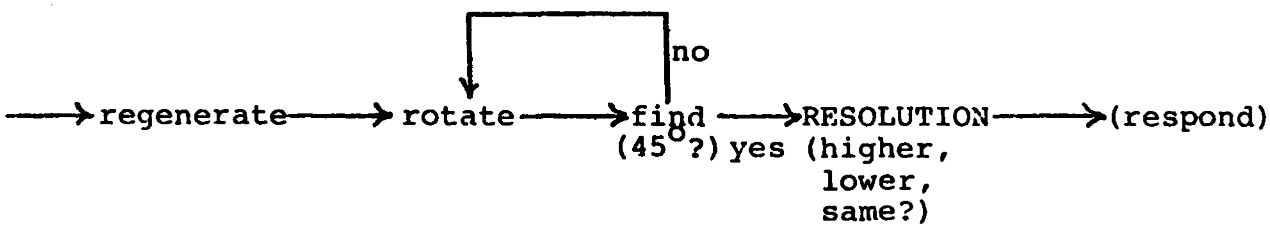
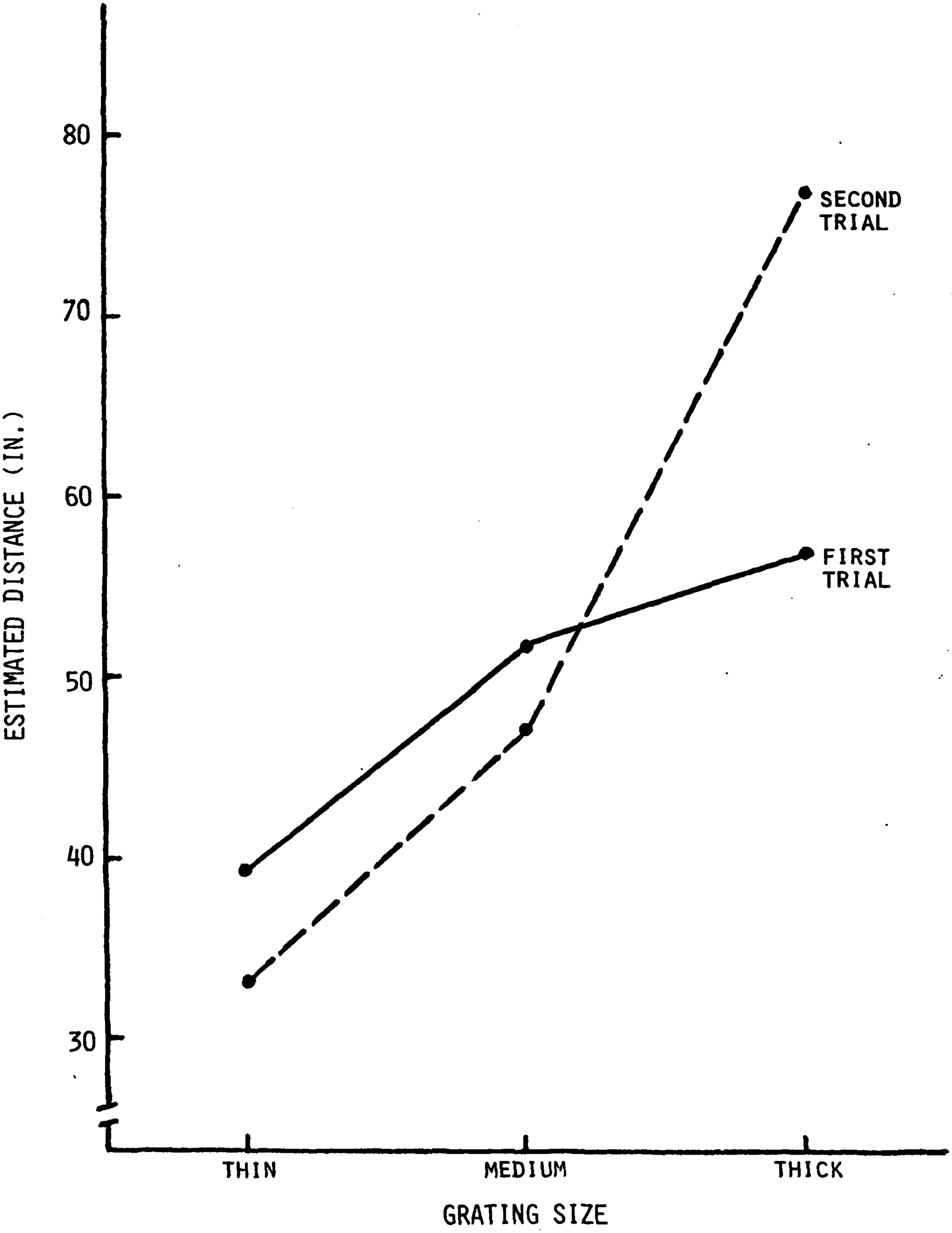


Figure 2. Model for oblique effect task

Figs 1 + 2



ESTIMATED DISTANCE (IN.)
80
70
60
50
40
30
SECOND TRIAL
FIRST TRIAL
THIN
MEDIUM
THICK
GRATING SIZE

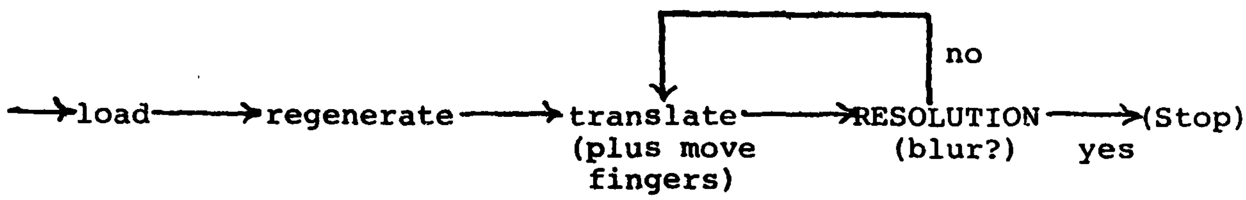
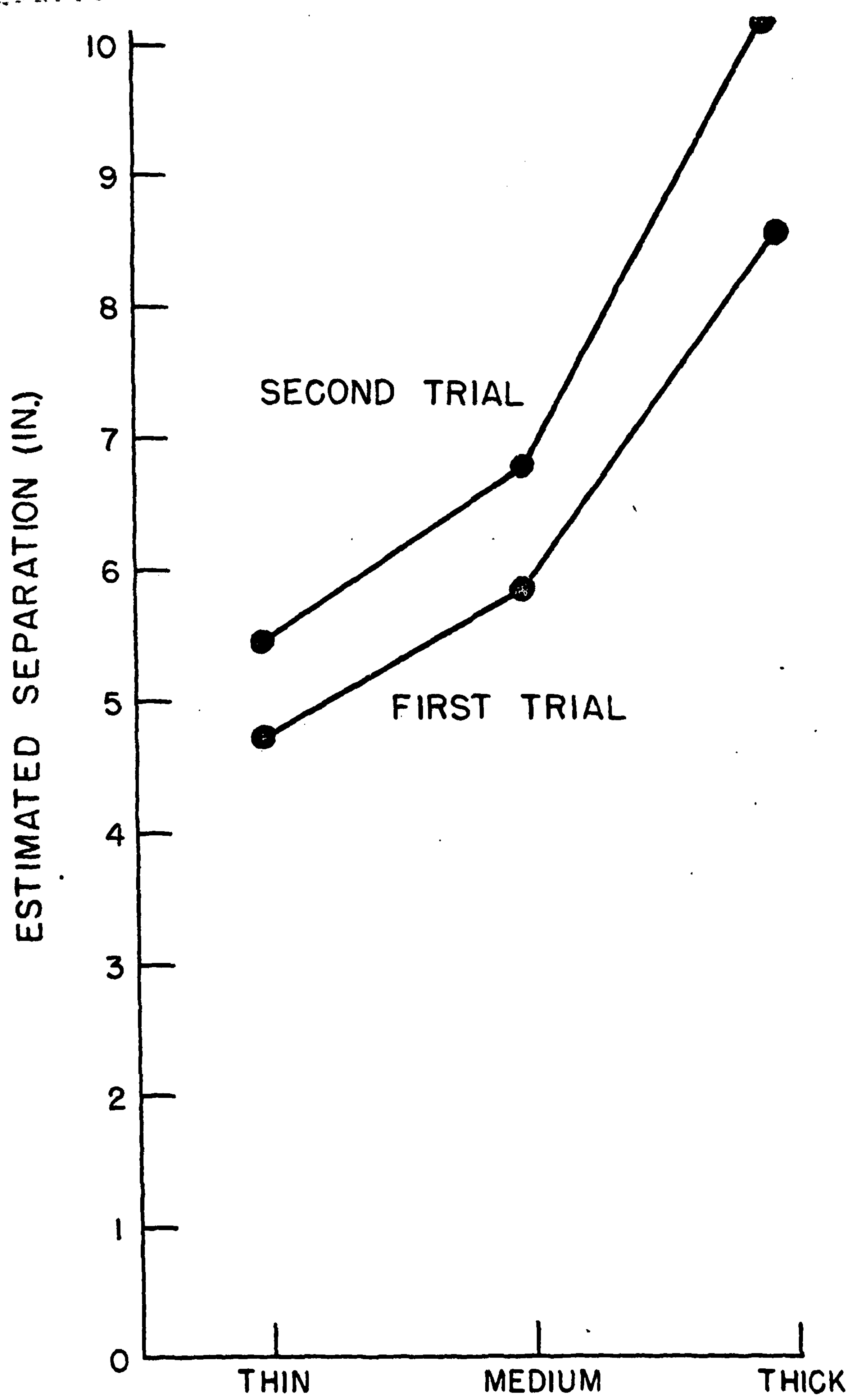


Figure 4. Model for the extent task.



SECOND TRIAL
FIRST TRIAL
ESTIMATED SEPARATION (IN.)
0
1
2
3
4
5
6
7
8
9
10
THIN
MEDIUM
THICK

Fig. 5

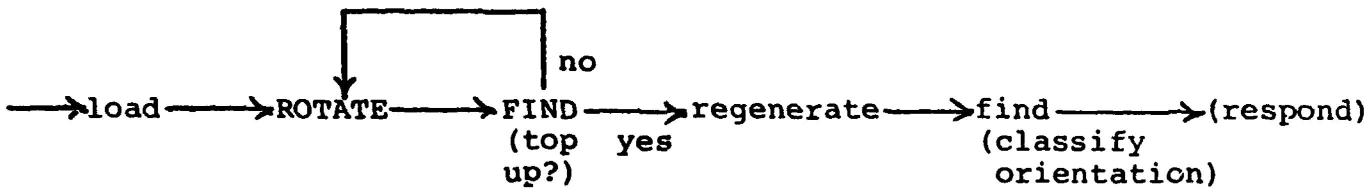
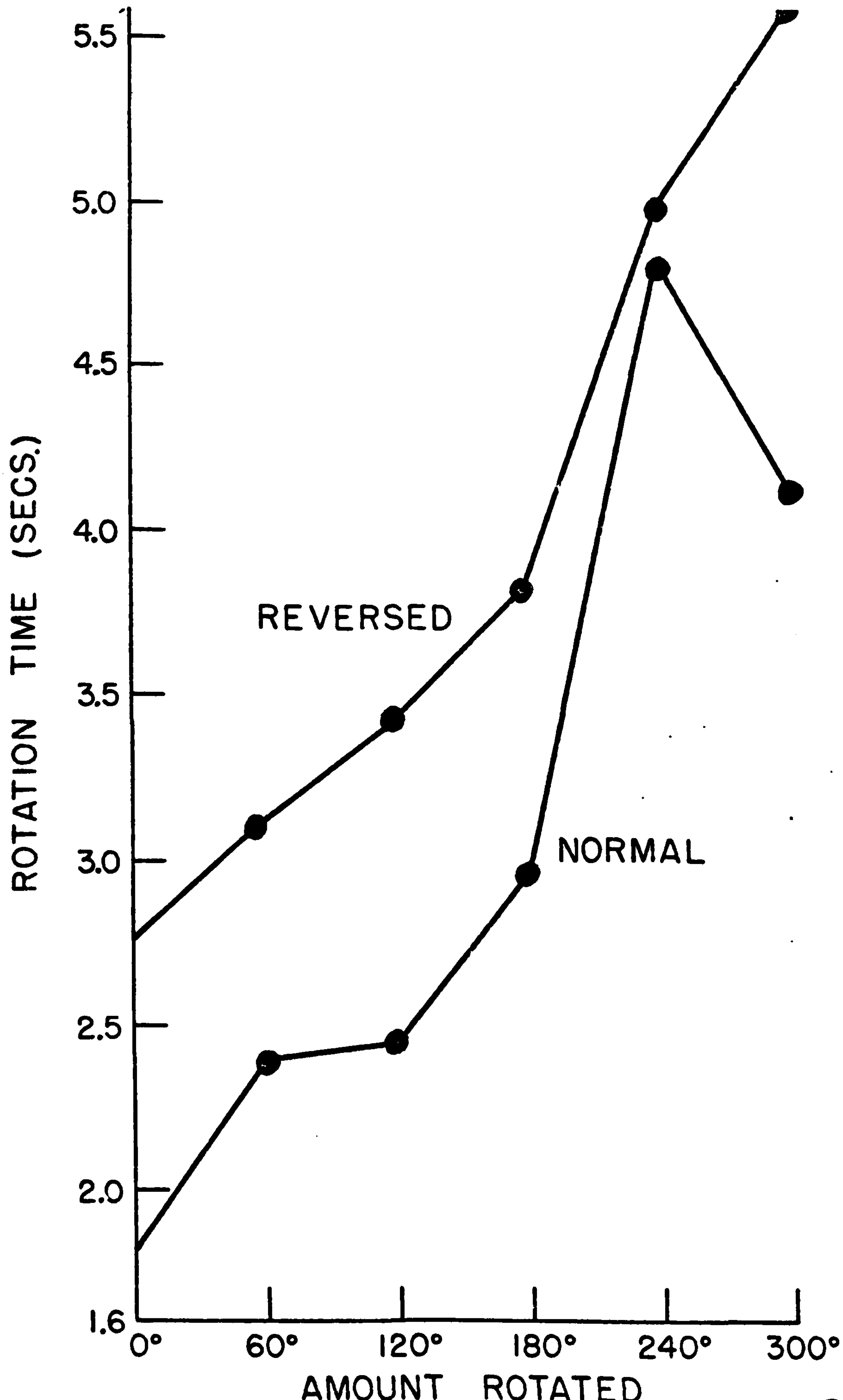
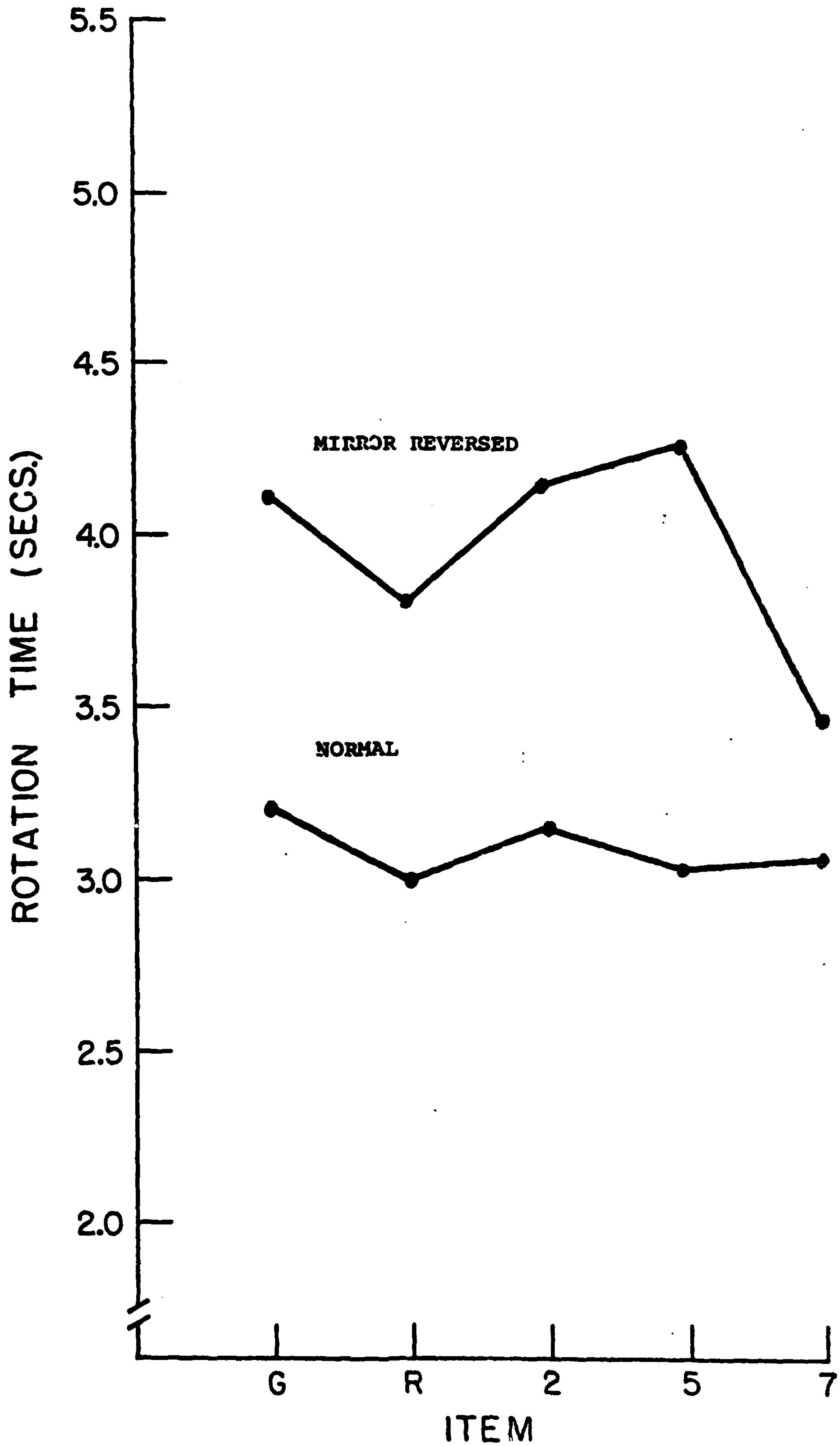


Figure 6. Model for the rotation task.



ROTATION TIME (SECS)
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.6
REVERSED
NORMAL
0°
60°
120°
180°
240°
300°
AMOUNT ROTATED

Fig 7



MIRROR REVERSED
NORMAL
ROTATION TIME (SECS.)
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
G
R
2
5
7
ITEM

Fig. 8

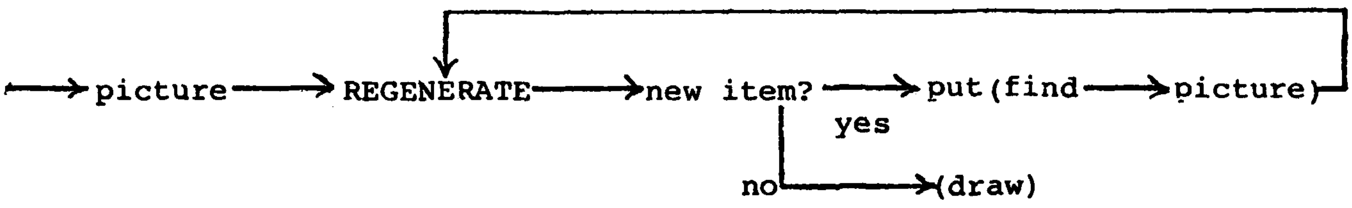


Figure 9. Model for line drawings task.

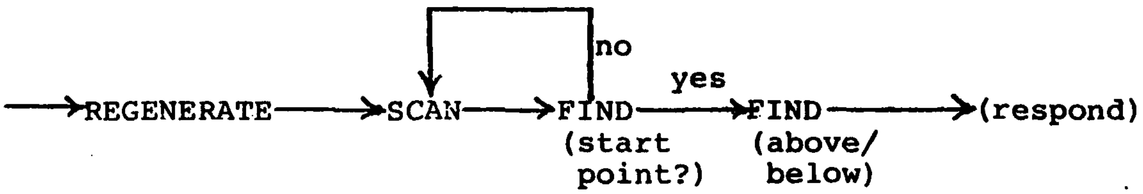


Figure 10. Model for line drawings judgment task.

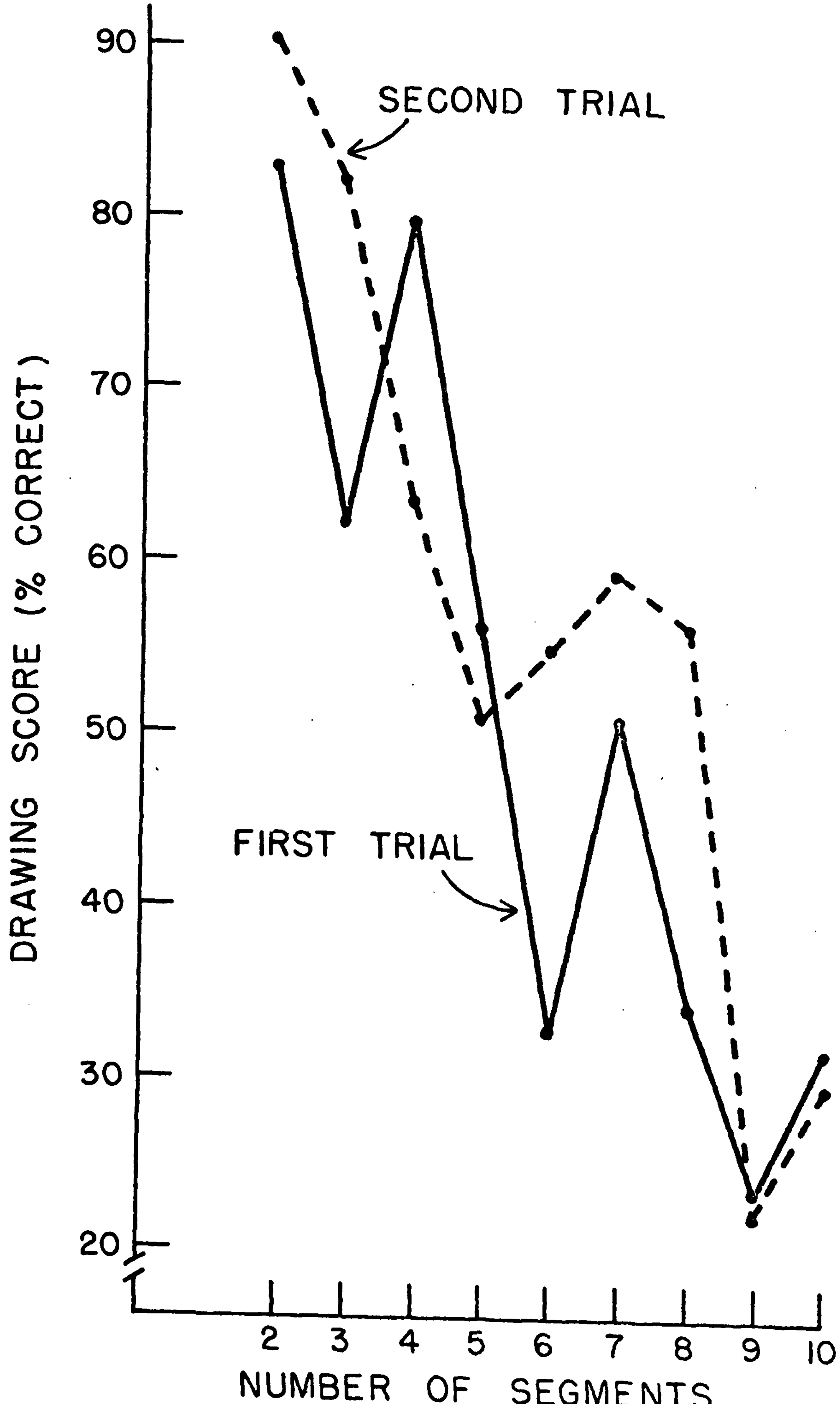
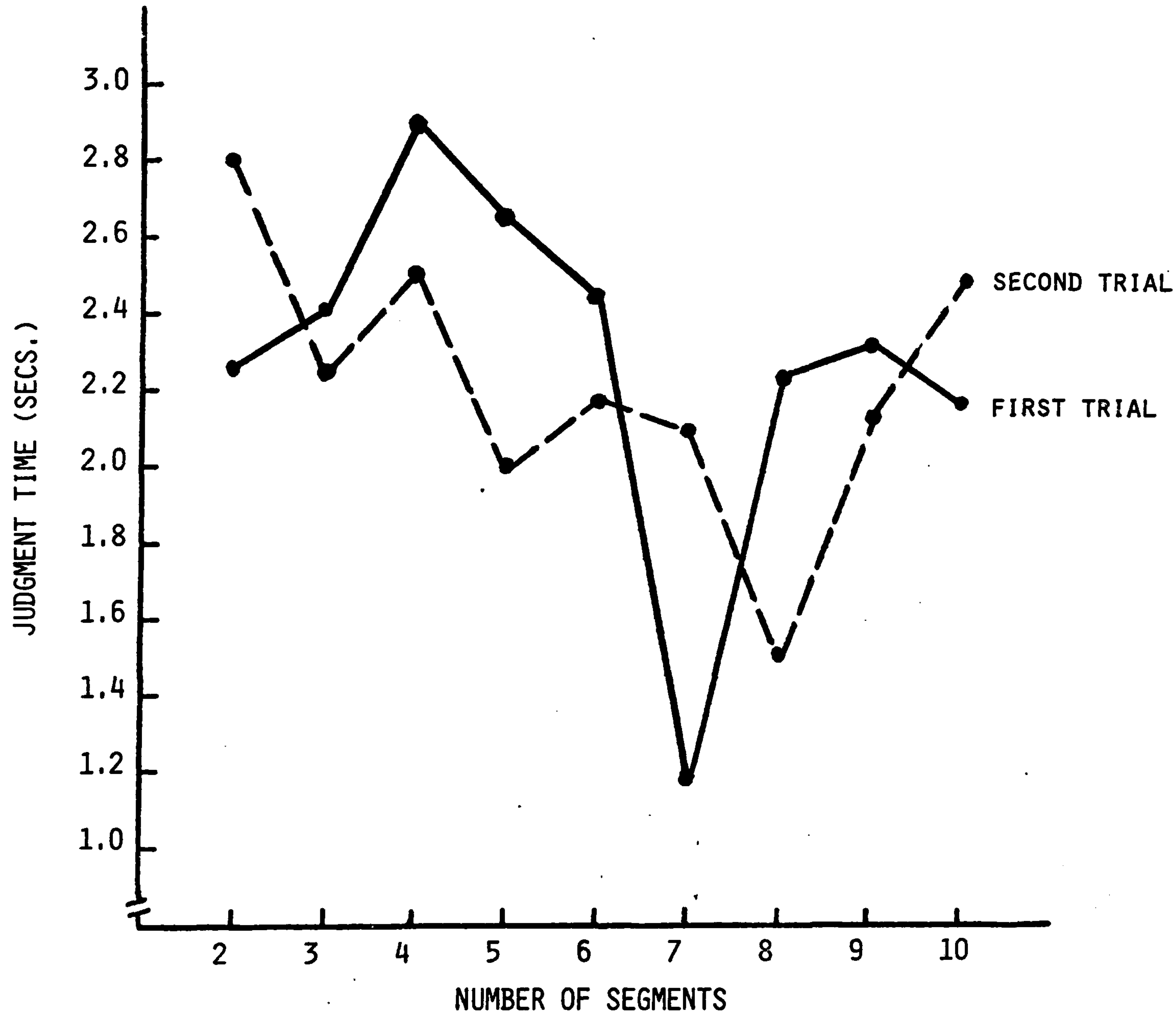

SECOND TRIAL
FIRST TRIAL
DRAWING SCORE (% CORRECT)
90
80
70
60
50
40
30
20
2
3
4
5
6
7
8
9
10
NUMBER OF SEGMENTS

Fig 11



JUDGMENT TIME (SECS.)
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
2
3
4
5
6
7
8
9
10
NUMBER OF SEGMENTS
SECOND TRIAL
FIRST TRIAL

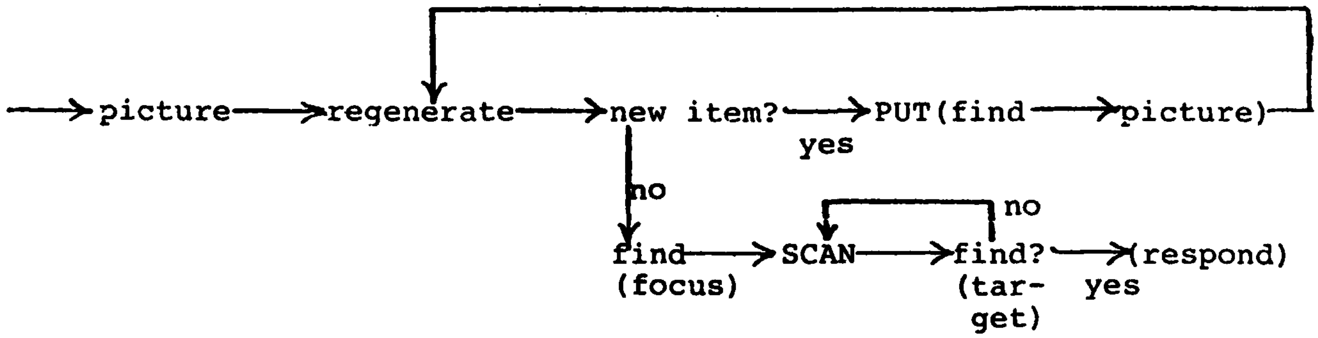
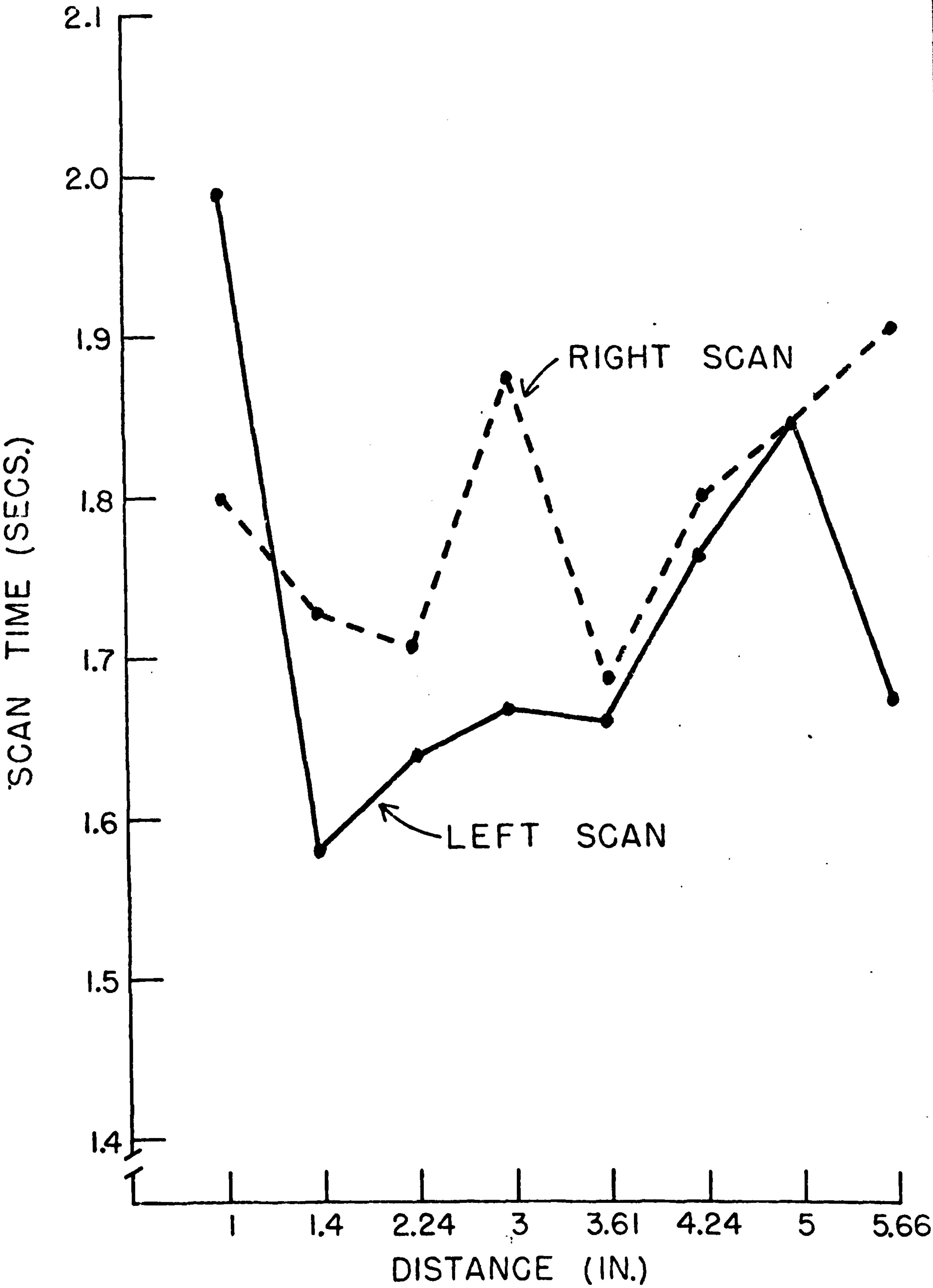


Figure 13. Model for the described scenes task.


SCAN TIME (SECS.)
2.1
2.0
1.9
1.8
1.7
1.6
1.5
1.4
RIGHT SCAN
LEFT SCAN
1
1.4
2.24
3
3.61
4.24
5
5.66
DISTANCE (IN.)

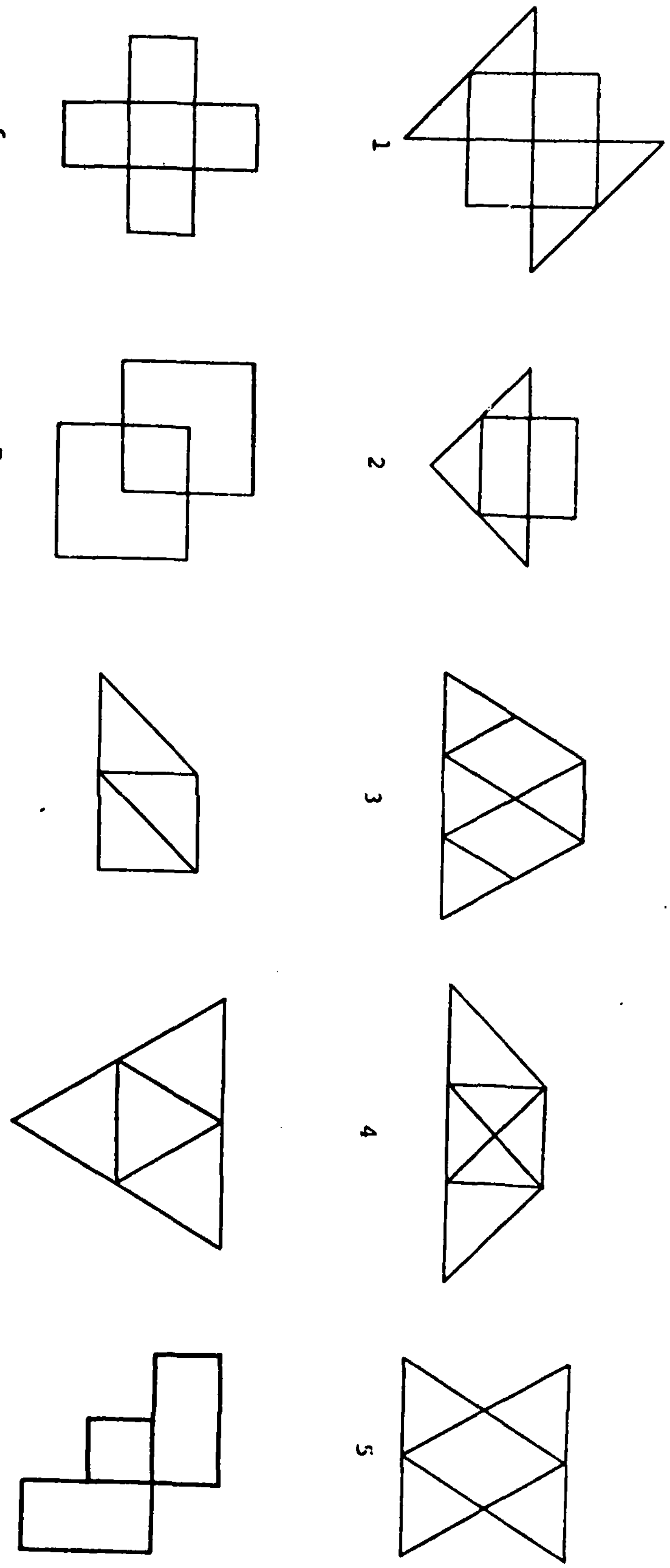
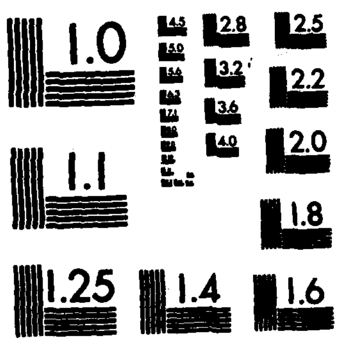


Fig 15

AD-A124 616 COMPONENTS OF MENTAL IMAGERY REPRESENTATION(U) BRANDEIS 2/2
UNIV WALTHAM MA S M KOSSLYN ET AL. 24 JAN 83 TR-1
N00014-79-C-0982

UNCLASSIFIED F/G 5/10 NL

END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

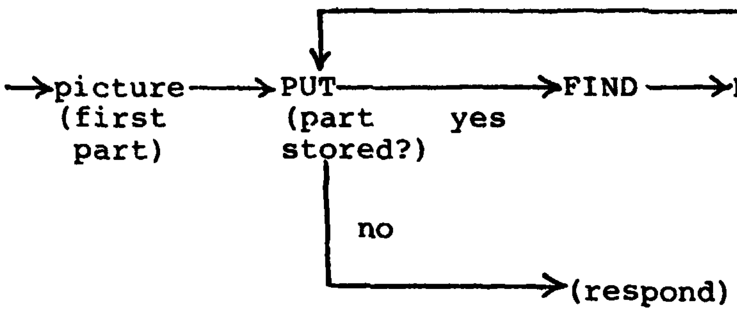


Figure 16. Model for the image generation task.

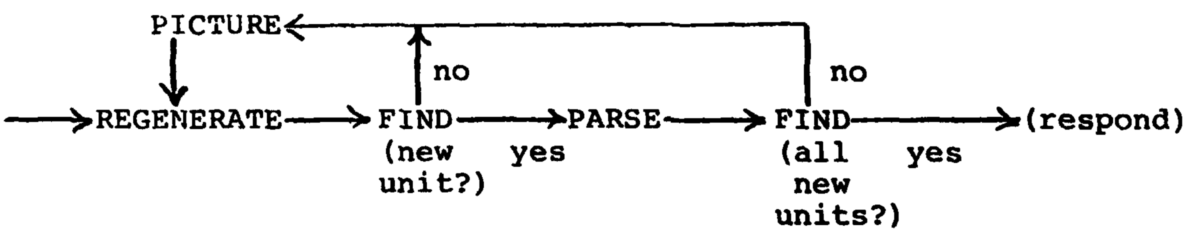


Figure 17. Model for the image reorganization task.

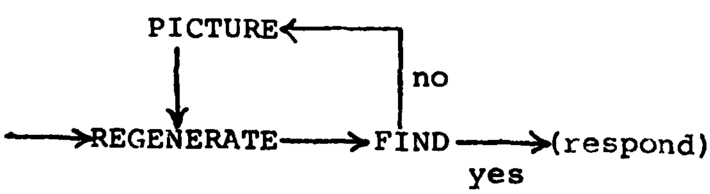
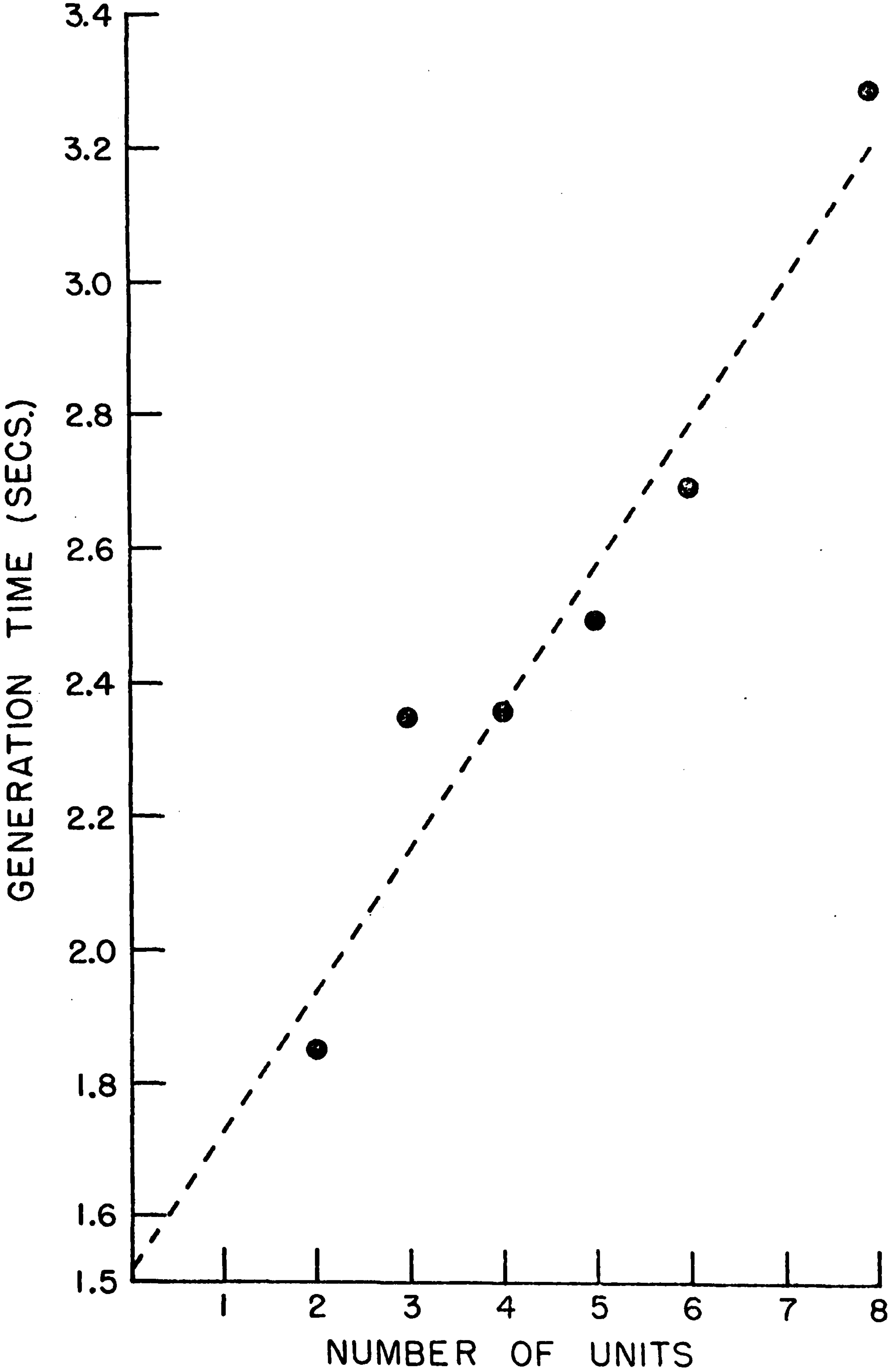
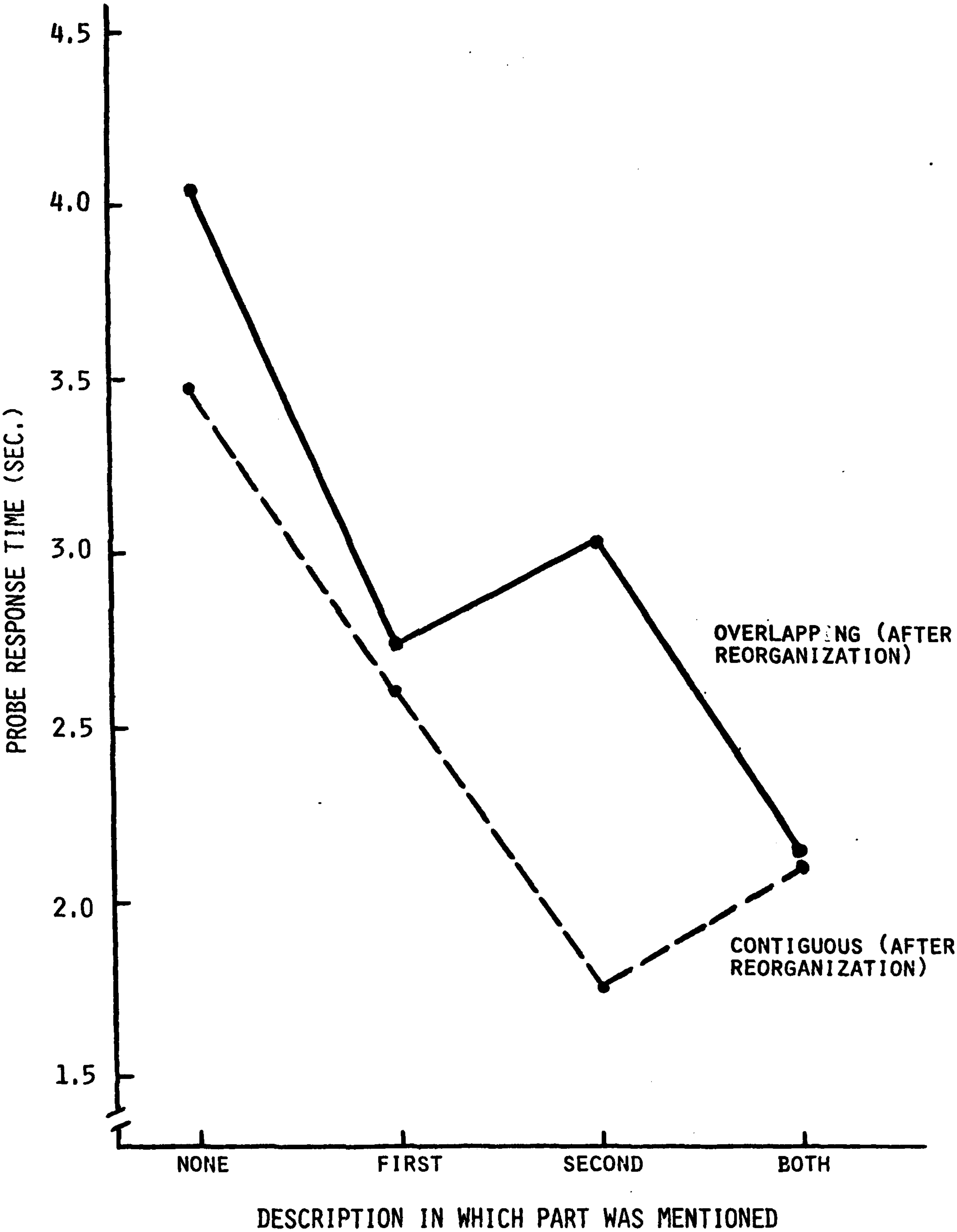


Figure 18. Model of the image inspection task.


GENERATION TIME (SECS.)
3.4
3.2
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.5
1
2
3
4
5
6
7
8
NUMBER OF UNITS


4.5
4.0
3.5
3.0
2.5
2.0
1.5
PROBE RESPONSE TIME (SEC.)
OVERLAPPING (AFTER REORGANIZATION)
CONTIGUOUS (AFTER REORGANIZATION)
NONE
FIRST
SECOND
BOTH
DESCRIPTION IN WHICH PART WAS MENTIONED

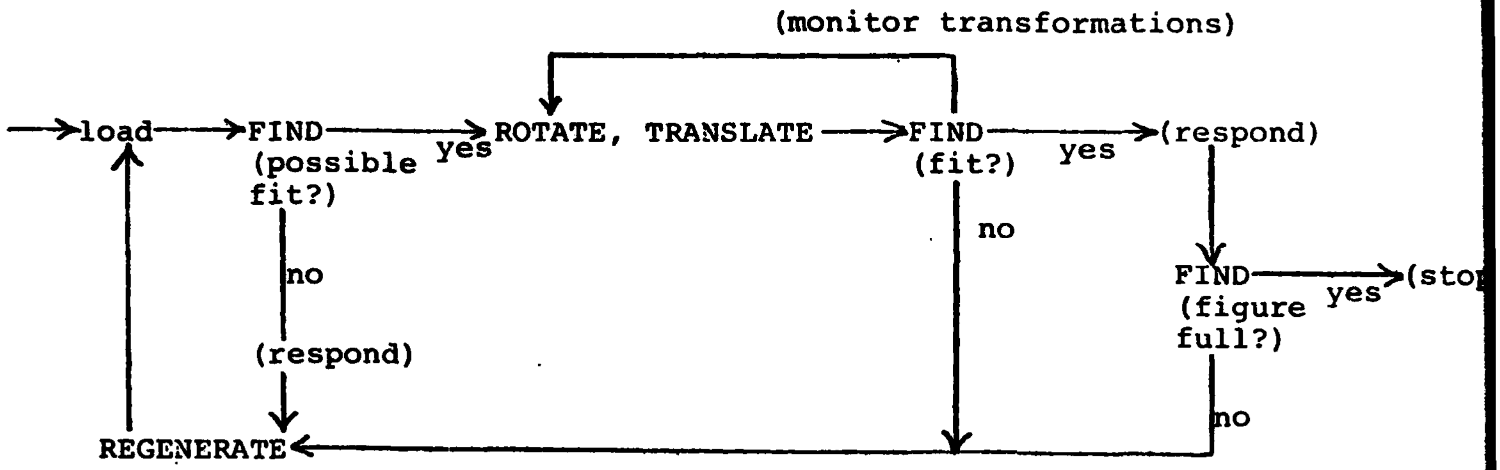


Figure 21. Model for Form Board test.

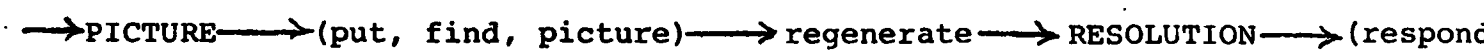


Figure 22. Model for VVIQ questionnaire task.

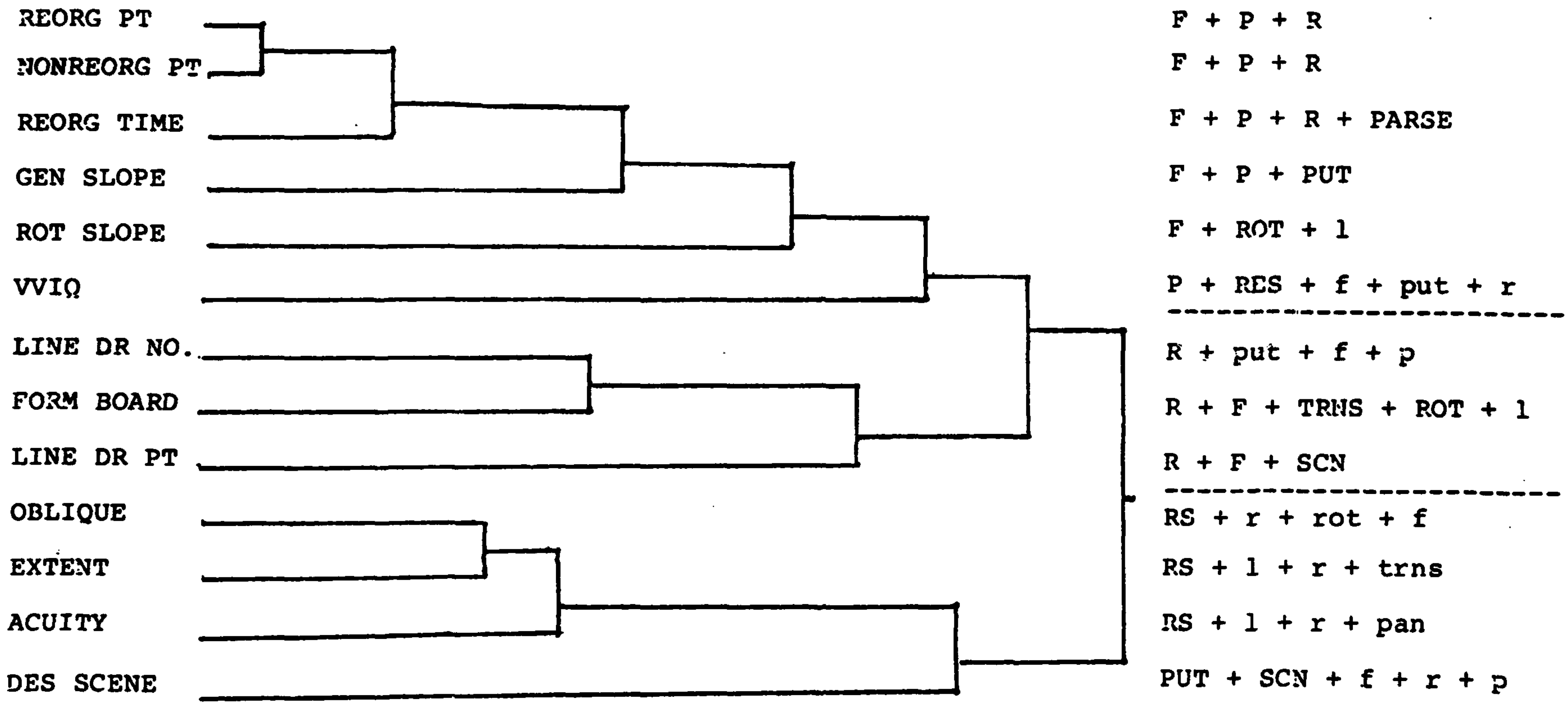

REORG PT
NONREORG PT
REORG TIME
GEN SLOPE
ROT SLOPE
VVIQ
LINE DR NO.
FORM BOARD
LINE DR PT
OBLIQUE
EXTENT
ACUITY
DES SCENE
F + P + R
F + P + R
F + P + R + PARSE
F + P + PUT
F + ROT + 1
P + RES + f + put + r
R + put + f + p
R + F + TRNS + ROT + 1
R + F + SCN
RS + r + rot + f
RS + 1 + r + trns
RS + 1 + r + pan
PUT + SCN + f + r + p

Table 1. Correlations among $\underline{z}$ scores for performance measures from the different tasks.[a]

		1	2	3	4	5	6	7	8	9	10	11	12	13
ACUITY	1	X												
OBLIQUE	2	.43	X											
EXTENT	3	.55	.59	X										
REORG PT	4	.00	.09	-.04	X									
NONREORG PT	5	.08	.19	.10	.79	X								
GEN SLOPE	6	.02	.26	.04	.30	.28	X							
REORG TIME	7	.02	.10	-.04	.64	.63	.26	X						
DES SCENE	8	-.08	-.13	-.17	-.27	-.25	-.10	-.18	X					
ROT SLOPE	9	-.09	.06	-.12	.21	.16	.02	.11	-.02	X				
LINE DR PT	10	.11	-.22	.16	.23	.01	.21	-.04	-.25	-.23	X			
LINE DR NO	11	-.19	-.10	.13	.28	.18	.05	.00	.05	-.09	.39	X		
FORM BOARD	12	-.31	.01	.02	.26	.32	-.04	-.01	.06	-.01	.01	.51	X	
VVIQ	13	-.06	-.44	-.26	.07	.06	-.19	.29	-.14	-.19	.08	.04	.07	X

Note a. $\underline{r}(49) = .28$, $p < .05$; $r(49) = .36$, $p < .01$

Table 2. Split-half reliabilities for each of the imagery tasks.[a,b]

Task	First Half	Second Half	r	t(first half - second half)[c]
ACUITY	49.1	52.9	.87	-1.27
OBLIQUE	.33	.30	.67	.73
EXTENT	6.4	7.6	.91	-3.80**
REORG PT	3119	2773	.86	-2.61*
NONREORG PT	2300	2507	.72	1.87
GEN SLOPE	337	181	.63	-2.54*
REORG TIME	2978	2266	.65	-3.61**
DES SCENE	.01	.05	.58	-.79
ROT SLOPE	10.1	12.4	.86	2.30*
LINE DR PT	2457	2067	.95	-5.35**
LINE DR NO	50.6	56.7	.80	-3.96**

Note a. Units are as described in the text.

Note b. The reliability used for the Form Board test was $\underline{r} = .81$, reported in Ekstrom et al., 1976 (pg 15); the reliability used for the VVIQ was the average of the .84 reported by Marks (1977) and the .82 obtained with our subjects.

Note c. Sigificant values here reveal the effects of practice (the sign of the value depends on the nature of the units of measurement); * = $p < .05$, ** = $p < .01$.

Table 3. Correlations among $\underline{z}$ scores from four non-imagery tests and the imagery performance measures

	MEM SPAN	VOCAB	DIFF USES	SYLLOGSM
ACUITY	.03	-.15	-.17	.06
OBLIQUE	-.02	-.01	-.08	.15
EXTENT	.14	-.04	-.09	.18
REORG PT	.35*	.11	-.01	.15
NONREORG PT	.17	.15	.06	.21
GEN SLOPE	-.02	.01	-.13	.16
REORG TIME	.18	.03	-.10	-.03
DES SCENE	-.07	-.09	-.03	-.18
ROT SLOPE	-.06	.10	-.08	.20
LINE DR PT	.24	.23	-.04	.35*
LINE DR NO	.45**	.37**	.29*	.35*
FORM BOARD	.19	.32*	.47**	.20
VVIQ	.02	.01	.25	-.03
MEM SPAN	X	.29*	-.05	.35*
VOCAB		X	.41**	.39**
DIFF USES			X	.28*
SYLLOGSM				X

Table 4. Components in models for the tasks.

	Components in Kosslyn & Shwartz Models[a]	Factors in Commonsense Theory[b]
ACUITY	RS + l + r + pan	v + c
OBLIQUE	RS + r + rot + f	m + v + c
EXTENT	RS + l + r + trns	v + c
REORG PT	F + P + R	s + m + v
NONREORG PT	F + P + R	s + m + v
GEN SLOPE	F + P + PUT	s
REORG TIME	F + PARSE + R + P	s + v + c
DES SCENE	PUT + SCN + f + r + p	m + c
ROT SLOPE	F + ROT + l	s + c
LINE DR PT	F + R + SCN	s + m + c
LINE DR NO	R + put + f + p	m
FORM BOARD	R + F + TRNS + ROT + l	s + m + v + c
VVIQ	RS + P + put + r + f	v

Note a. Capitol letters in the Kosslyn & Shwartz models indicate that a component was weighted. The abbreviations used are as follows: rs: RESOLUTION; l: LOAD; r:REGENERATE: pan: PAN; rot: ROTATE; f: FIND; trns: TRANSLATE; p:PICTURE; put: PUT; parse: PARSE; scn:SCAN.

Note b. The abbreviations used are as follows: v:vividness; m: memory capacity; s: speed; c: control (of transformations).

Table 5. Predicted similarities among task performance, using the Kosslyn & Shwartz theory.

		1	2	3	4	5	6	7	8	9	10	11	12	13
ACUITY	1	X												
OBLIQUE	2	0.67	X											
EXTENT	3	1.00	0.67	X										
REORG PT	4	0.08	0.22	0.08	X									
NONREORG PT	5	0.08	0.22	0.08	3.00	X								
GEN SLOPE	6	0.00	0.08	0.00	0.75	0.75	X							
REORG TIME	7	0.07	0.17	0.07	1.50	1.50	0.55	X						
DES SCENE	8	0.07	0.15	0.07	0.33	0.33	0.71	0.25	X					
ROT SLOPE	9	0.08	0.29	0.08	0.27	0.27	0.27	0.21	0.08	X				
LINE DR PT	10	0.08	0.22	0.08	0.75	0.75	0.23	0.55	0.71	0.27	X			
LINE DR NO	11	0.11	0.25	0.11	1.25	1.25	0.50	0.71	0.57	0.10	0.57	X		
FORM BOARD	12	0.13	0.30	0.25	0.50	0.50	0.18	0.40	0.13	0.78	0.50	0.36	X	
VVIQ	13	0.44	0.63	0.44	0.71	0.71	0.71	0.50	0.40	0.08	0.17	0.57	0.13	X

Table 6. Regression analysis using no weights.

Variable	Multiple r	Beta	Predicted Weight
RESOLUTION	.270	.28139	1.0
SCAN	.388	.32299	1.0
PUT	.434	.18913	.5
ROTATE	.457	.13787	.67
REGENERATE	.474	.14769	.55
FIND	.485	.14379	.64
PICTURE	.487	.08099	.71
LOAD	.490	-.07404	0
PARSE	.491	-.01869	-

Table 7. Intercorrelations among rotated factors.

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
FACTOR 1	1.000				
FACTOR 2	0.020	1.000			
FACTOR 3	0.016	-0.076	1.000		
FACTOR 4	0.091	0.073	0.007	1.000	
FACTOR 5	0.116	0.092	0.053	-0.100	1.000

Table 8. Rotated factor loadings (pattern matrix).

		FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
ACUITY	1	0.066	0.754	-0.287	0.098	-0.144
OBLIQUE	2	0.108	0.784	0.018	-0.271	0.284
EXTENT	3	-0.049	0.891	0.170	0.118	-0.002
REORG PT	4	0.839	-0.045	0.228	0.081	0.155
NON REORG PT	5	0.859	0.140	0.257	-0.120	0.036
GEN SLOPE	6	0.250	-0.053	-0.131	0.327	0.713
REORG TIME	7	0.876	-0.037	-0.116	-0.104	-0.096
DES SCENE	8	-0.397	-0.150	0.203	-0.343	0.052
ROT SLOPE	9	0.222	-0.163	-0.066	-0.476	0.365
LINE DR PT	10	-0.019	-0.050	0.112	0.913	0.168
LINE DR NO	11	0.031	-0.024	0.797	0.366	0.040
FORM BOARD	12	0.155	-0.002	0.875	-0.130	-0.116
VVIQ	13	0.359	-0.287	-0.072	0.163	-0.743

Navy

1 Robert Ahlers
Code N711
Human Factors Laboratory
NAVTRAEQUIPCEN
Orlando, FL 32813

1 Dr. Arthur Bachrach
Environmental Stress Program Center
Naval Medical Research Institute
Bethesda, MD 20014

1 Dr. Meryl S. Baker
Navy Personnel R&D Center
San Diego, CA 92152

1 CDR Robert J. Biersner
Naval Medical R&D Command
National Naval Medical Center
Bethesda, MD 20814

1 Dr. Alvah Bittner
Naval Biodynamics Laboratory
New Orleans, LA 70189

1 Liaison Scientist
Office of Naval Research
Branch Office, London
Box 39
FPO New York, NY 09510

1 Dr. Stanley Collyer
Office of Naval Technology
800 N. Quincy Street
Arlington, VA 22217

1 CDR Mike Curran
Office of Naval Research
800 N. Quincy St.
Code 270
Arlington, VA 22217

1 Dr. Tom Duffy
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Carl E. Englund
Naval Health Research Center
Code 8060 Environmental Physiology Dept
P.O. Box 85122
San Diego, CA 92138

Navy

1 DR. PAT FEDERICO
Code P13
NPRDC
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 LT Steven D. Harris, MSC, USN
RFD 1, Box 243
Riner, VA 24149

1 Dr. Jim Hollan
Code 304
Navy Personnel R & D Center
San Diego, CA 92152

1 Dr. Ed Hutchins
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. Peter Kincaid
Training Analysis & Evaluation Group
Dept. of the Navy
Orlando, FL 32813

1 Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. James Lester
ONR Detachment
495 Summer Street
Boston, MA 02210

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 Dr. James McBride
Navy Personnel R&D Center
San Diego, CA 92152

Navy

1 Dr. George Moeller
Director, Behavioral Sciences Dept.
Naval Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 63409

1 Dr William Montague
NPRDC Code 13
San Diego, CA 92152

1 Naval Ocean R&D Agency
NSTL Station
Attn: LCDR J. D. McKendrick
Code 335
Bay St. Louis, MO 39529

1 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

1 Office of Naval Research
Code 433
800 N. Quincy SStreet
Arlington, VA 22217

1 Office of Naval Research
Code 441NP
800 N. Quincy Street
Arlington, VA 22217

6 Personnel & Training Research Group
Code 442PT
Office of Naval Research
Arlington, VA 22217

1 Office of the Chief of Naval Operations
Research Development & Studies Branch
OP 115
Washington, DC 20350

1 Daira Paulson
Code 14 - Training Systems
Navy Personel R&D Center
San Diego, CA 92152

1 LT Frank C. Petho, MSC, USN (Ph.D)
CNET (N-432)
NAS
Pensacola, FL 32508

Navy

1 Dr. Gary Poock
Operations Research Department
Code 55PK
Naval Postgraduate School
Monterey, CA 93940

1 Dr. Bernard Rimland (01C)
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Worth Scanland
CNET (N-5)
NAS, Pensacola, FL 32508

1 Mr. Irving Schiff
Dept. of the Navy
Chief of Naval Operations
OP 113
Washington, DC 20350

1 Dr. Sam Schiflett
Aircrew Systems (SY70E)
Naval Air Test Center
Patuxent River, MD 20670

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Alfred F. Smode, Director
Training Analysis & Evaluation Group
Dept. of the Navy
Orlando, FL 32813

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Frederick Steinheiser
CNO - OP115
Navy Annex
Arlington, VA 20370

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

1 Roger Weissinger-Baylon
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940

Navy

1 Dr. Ronald Weitzman
Code 54 WZ
Department of Administrative Sciences
U. S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Douglas Wetzel
Code 12
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152

Marine Corps

1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

Army

1 Technical Director
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Beatrice J. Farr
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Marshall Narva
US Army Research Institute for the
Behavioral & Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Harold F. O'Neil, Jr.
Director, Training Research Lab
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Psotka
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Sasmor
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Wisher
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 AFHRL/LRS
Attn: Susan Ewing
WPAFB
WPAFB, OH 45433

1 U.S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, DC 20332

1 Air University Library
AUL/LSE 76/443
Maxwell AFB, AL 36112

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235

1 Dr. Alfred R. Fregly
AFOSR/NL
Bolling AFB, DC 20332

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Dr. David R. Hunter
AFHRL/MO
Brooks AFB, TX 78235

1 Dr. T. M. Longridge
AFHRL/OTGT
Williams AFB, AZ 85224

1 Dr. Joseph Yasatuke
AFHRL/OT
Williams AFB, AZ 58224

Department of Defense

12 Defense Technical Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC

1 Military Assistant for Training and
Personnel Technology
Office of the Under Secretary of Defens
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 Major Jack Thorpe
DARPA
1400 Wilson Blvd.
Arlington, VA 22209

Civilian Agencies

1 Dr. Patricia A. Butler
NIE-BRN Bldg, Stop # 7
1200 19th St., NW
Washington, DC 20208

1 Dr. Paul G. Chapin
Linguistics Program
National Science Foundation
Washington, DC 20550

1 Dr. Susan Chipman
Learning and Development
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Arthur Melmed
OERI
1200 19th Street NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Office of Scientific and Engineering
Personnel and Education
National Science Foundation
Washington, DC 20550

1 Dr. Judith Orasanu
National Institute of Education
1200 19th St., N.W.
Washington, DC 20208

1 Dr. Ramsay W. Selden
National Institute of Education
1200 19th St., NW
Washington, DC 20208

1 Chief, Psychological Reserch Branch
U. S. Coast Guard (G-P-1/2/TP42)
Washington, DC 20593

1 Dr. Frank Withrow
U. S. Office of Education
400 Maryland Ave. SW
Washington, DC 20202

Civilian Agencies

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Private Sector

1 Dr. John R. Anderson
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Dr. John Annett
Department of Psychology
University of Warwick
Coventry CV4 7AJ
ENGLAND

1 Dr. Michael Atwood
Bell Laboratories
11900 North Pecos St.
Denver, CO 80234

1 Psychological Research Unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600
AUSTRALIA

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

1 Dr. Patricia Baggett
Department of Psychology
University of Colorado
Boulder, CO 80309

1 Dr. Jonathan Baron
80 Glenn Avenue
Berwyn, PA 19312

1 Dr. Jackson Beatty
Department of Psychology
University of California
Los Angeles, CA 90024

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

1 Dr. John Black
Yale University
Box 11A, Yale Station
New Haven, CT 06520

Private Sector

1 Dr. R. Darrel Bock
Department of Education
University of Chicago
Chicago, IL 60637

1 Dr. Lyle Bourne
Department of Psychology
University of Colorado
Boulder, CO 80309

1 Dr. John S. Brown
XEROX Palo Alto Research Center
3333 Coyote Road
Palo Alto, CA 94304

1 Bundministerium der Verteidigung
-Referat P II 4-
Psychological Service
Postfach 1328
D-5300 Bonn 1
F. R. of Germany

1 Dr. Pat Carpenter
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Dr. John B. Carroll
409 Elliott Rd.
Chapel Hill, NC 27514

1 Dr. William Chase
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Micheline Chi
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. Michael Cole
University of California
at San Diego
Laboratory of Comparative
Human Cognition - D003A
La Jolla, CA 92093

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

Private Sector

1 Dr. Lynn A. Cooper
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. Kenneth B. Cross
Anacapa Sciences, Inc.
P.O. Drawer Q
Santa Barbara, CA 93102

1 Dr. Diane Damos
Arizona State University
Department of Psychology
Tempe, AZ 85287

1 LCOL J. C. Eggenberger
DIRECTORATE OF PERSONNEL APPLIED RESEAR
NATIONAL DEFENCE HQ
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A

1 Dr. Jeffrey Elman
University of California, San Diego
Department of Linguistics
La Jolla, CA 92093

1 Dr. Paul Feltovich
Department of Medical Education
Southern Illinois University
School of Medicine
P.O. Box 3926
Springfield, IL 62708

1 Professor Reuven Feuerstein
HWCRI Rehov Karmon 6
Bet Hakerem
Jerusalem
Israel

1 Mr. Wallace Feurzeig
Department of Educational Technology
Bolt Beranek & Newman
10 Moulton St.
Cambridge, MA 02238

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

Private Sector

1 Univ. Prof. Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

1 Dr. Dexter Fletcher
WICAT Research Institute
1875 S. State St.
Orem, UT 22333

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Alinda Friedman
Department of Psychology
University of Alberta
Edmonton, Alberta
CANADA T6G 2E9

1 Dr. Robert Glaser
Learning Research & Development Center
University of Pittsburgh
3939 O'Hara Street
PITTSBURGH, PA 15260

1 Dr. Marvin D. Glock
217 Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Josph Goguen
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025

1 Dr. Daniel Gopher
Department of Psychology
University of Illinois
Champaign, IL 61820

1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Harold Hawkins
Department of Psychology
University of Oregon
Eugene, OR 97403

Private Sector

1 Dr. Barbara Hayes-Roth
Department of Computer Science
Stanford University
Stanford, CA 95305

1 Dr. Frederick Hayes-Roth
Teknowledge
525 University Ave.
Palo Alto, CA 94301

1 Dr. James R. Hoffman
Department of Psychology
University of Delaware
Newark, DE 19711

1 Dr. Kristina Hooper
Clark Kerr Hall
University of California
Santa Cruz, CA 95060

1 Glenda Greenwald, Ed.
Human Intelligence Newsletter
P. O. Box 1163
Birmingham, MI 48012

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Steven W. Keele
Dept. of Psychology
University of Oregon
Eugene, OR 97403

1 Dr. Scott Kelso
Haskins Laboratories, Inc
270 Crown Street
New Haven, CT 06510

1 Dr. David Kieras
Department of Psychology
University of Arizona
Tuscon, AZ 85721

1 Dr. Walter Kintsch
Department of Psychology
University of Colorado
Boulder, CO 80302

Private Sector

1 Dr. Pat Langley
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Dr. Marcy Lansman
The L. L. Thurstone Psychometric Laboratory
University of North Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Dr. Jill Larkin
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

1 Dr. Jim Levin
University of California at San Diego
Laboratory fof Comparative Human Cognition - D003A
La Jolla, CA 92093

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Dr. Jay McClelland
Department of Psychology
MIT
Cambridge, MA 02139

1 Dr. James R. Miller
Texas Instruments, Inc.
Central Research Laboratory
P. O. Box 226015, MS238
Dallas, TX 75266

1 Dr. Allen Munro
Behavioral Technology Laboratories
1845 Elena Ave., Fourth Floor
Redondo Beach, CA 90277

1 Dr. Donald A Norman
Cognitive Science, C-015
Univ. of California, San Diego
La Jolla, CA 92093

Private Sector

1 Committee on Human Factors
JH 811
2101 Constitution Ave. NW
Washington, DC 20418

1 Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 Dr. James W. Pellegrino
University of California, Santa Barbara
Dept. of Psychology
Santa Barabara, CA 93106

1 Dr. Nancy Pennington
University of Chicago
5801 S. Ellis Avenue
Chicago, IL 60637

1 Mr. L. Petrullo
2431 N. Edgewood Street
ARLINGTON, VA 22207

1 Dr. Martha Polson
Department of Psychology
Campus Box 346
University of Colorado
Boulder, CO 80309

1 DR. PETER POLSON
DEPT. OF PSYCHOLOGY
UNIVERSITY OF COLORADO
BOULDER, CO 80309

1 Dr. Steven E. Poltrock
Department of Psychology
University of Denver
Denver, CO 80208

1 Dr. Mike Posner
Department of Psychology
University of Oregon
Eugene, OR 97403

Private Sector

1 Dr. Fred Reif
Physics Department
University of California
Berkeley, CA 94720

1 Dr. Lauren Resnick
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 1521

1 Mary S. Riley
Program in Cognitive Science
Center for Human Information Processing
University of California, San Diego
La Jolla, CA 92093

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
Murray Hill, NJ 07974

1 Dr. William B. Rouse
Georgia Institute of Technology
School of Industrial & Systems
Engineering
Atlanta, GA 30332

1 Dr. David Rumelhart
Center for Human Information Processing
Univ. of California, San Diego
La Jolla, CA 92093

1 Dr. Michael J. Samet
Perceptronics, Inc
6271 Variel Avenue
Woodland Hills, CA 91364

1 Dr. Arthur Samuel
Yale University
Department of Psychology
Box 11A, Yale Station
New Haven, CT 06520

1 Dr. Walter Schneider
Psychology Department
603 E. Daniel
Champaign, IL 61820

Private Sector

1 Dr. Alan Schoenfeld
Mathematics and Education
The University of Rochester
Rochester, NY 14627

1 DR. ROBERT J. SEIDEL
INSTRUCTIONAL TECHNOLOGY GROUP
HUMRRO
300 N. WASHINGTON ST.
ALEXANDRIA, VA 22314

1 Mr. Colin Sheppard
Applied Psychology Unit
Admiralty Marine Technology Est.
Teddington, Middlesex
United Kingdom

1 Dr. David Shucard
Brain Sciences Labs
National Jewish Hospital Research Cente
National Asthma Center
Denver, CO 80206

1 Robert S. Siegler
Associate Professor
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh, PA 15213

1 Dr. H. Wallace Sinaiko
Program Director
Manpower Research and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

1 Dr. Edward E. Smith
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305

1 Dr. Eliott Soloway
Yale University
Department of Computer Science
P.O. Box 2158
New Haven, CT 06520

Private Sector

1 Dr. Kathryn T. Spoehr
Psychology Department
Brown University
Providence, RI 02912

1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 Dr. Albert Stevens
Bolt Beranek & Newman, Inc.
10 Moulton St.
Cambridge, MA 02238

1 David E. Stone, Ph.D.
Hazeltine Corporation
7680 Old Springhouse Road
McLean, VA 22102

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

1 Dr. Kikumi Tatsuoka
Computer Based Education Research Lab
252 Engineering Research Laboratory
Urbana, IL 61801

1 Dr. Perry W. Thorndyke
Perceptronics, Inc.
545 Middlefield Road, Suite 140
Menlo Park, CA 94025

1 Dr. Douglas Towne
Univ. of So. California
Behavioral Technology Labs
1845 S. Elena Ave.
Redondo Beach, CA 90277

1 Dr. Don Weitzman
Bell Laboratories
Room 2A-202
Whippany Road
Whippany, NJ 07931

1 Dr. Keith T. Wescourt
Perceptronics, Inc.
545 Middlefield Road, Suite 140
Menlo Park, CA 94025

Private Sector

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
Lawrence, KS 66045

1 William B. Whitten
Bell Laboratories
2D-610
Holmdel, NJ 07733

1 Dr. Christopher Wickens
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Dr. Mike Williams
Zerox PARC
3333 Coyote Hill Road
Palo Alto, CA 94304

3-8

DT